

10/593119  
IAP9/Rec'd PCT/PTO 18 SEP 2005**Novel alcohol dehydrogenases**

The invention relates to novel polypeptides which have the biological activity of an NAD- or NADP-dependent alcohol dehydrogenase. The invention furthermore relates to nucleic acids encoding said polypeptides, to nonhuman hosts or host cells and to reaction systems which may be used for preparing desired products. The polypeptides of the invention are preferably used in the preparation, starting from aldehydes or ketones, of primary and enantiomerically pure secondary alcohols which may serve as intermediates for medicaments. Alternatively, the polypeptides of the invention may also be employed in the reverse reaction, i.e. the oxidation of alcohols with the formation of aldehydes or ketones.

15

The description makes reference to a number of documents. The disclosure content of these documents is hereby incorporated by reference.

20 Enantiomerically pure alcohols are among the most important chiral building blocks of industrial special and fine chemistry. These products act, inter alia, as essential key intermediates in the preparation of medicaments. For a long time, the industrial route to these target molecules went preferably via purely chemical processes, for example by way of resolution of the racemate. This involves, starting from a ketone, firstly preparing the alcohol in its racemic form and then isolating the desired enantiomer in a resolution of the racemate with the aid of at least stoichiometric amounts of a chiral auxiliary substance. Disadvantages of these methods include not only the 50% maximum yield of the resolution of the racemate but must also be seen in the use of numerous ecologically problematic starting compounds for preparing the racemate. Further disadvantages are the additional step of recycling the undesired enantiomer as well as the need of chiral auxiliary reagents (moreover, stoichiometric amounts

**BEST AVAILABLE COPY**

thereof) for the resolution of the racemate. The concept of the resolution of the racemate is summarized in the equation (1a) in the overview in Figure 1. A first substantial progress toward a more sustainable process was achieved using the biocatalytic resolution of the racemate, thereby dispensing with the necessity of employing stoichiometric amounts of chiral auxiliary reagents. Regrettably, however, all other disadvantages listed above remained relevant, despite such a biocatalytic route.

10

One possible way of avoiding the above-described disadvantages of the resolution of the racemate or of diastereoselective syntheses is the direct conversion of ketones to the desired optically active alcohols in one step. Such "direct asymmetrical processes" may be carried out firstly by using metal-containing chemocatalysts, with the chemocatalysts employed being heavy metal-containing complexes which include a chiral ligand. Besides the use of ecologically problematic heavy metals as a substantial catalyst component, the need for expensive and partly very sensitive ligands, for example phosphane ligands, is also disadvantageous.

20

Another alternative is the direct asymmetrical reduction using suitable biocatalysts for quantitative conversion of prochiral substrates to the desired enantiomerically pure product. Here too, the number of reaction steps is reduced to the theoretically possible minimum of only one step, the biocatalytic conversion is carried out under ecologically excellent conditions (inter alia water as a solvent), and the process as such proceeds with high atom economy. The concept of a biocatalytic and sustainable process of this kind is set out in the equation (1b) of the overview in Figure 1.

30

35

One disadvantage of the biocatalytic variant, however, is the lack of alcohol dehydrogenases available on an

industrial scale as suitable biocatalysts for the target reaction and expression thereof. The object on which the present invention is based was therefore to obtain novel, efficient and industrially usable alcohol dehydrogenases.

5 This object is achieved by the embodiments characterized in the claims.

Thus, the invention relates to a polypeptide which has the biological activity of an NAD- or NADP-dependent alcohol

10 dehydrogenase and which comprises or has one of the following sequences: the sequence of SEQ ID NO.: 1, the sequence of SEQ ID NO.: 2, the sequence of SEQ ID NO.: 3 or a sequence which is at least 90% identical to the sequence of SEQ ID NO.: 3. Said sequence is preferably at least 95%

15 identical to SEQ ID NO.: 3. More preferably, said sequence is at least 98% or 99% identical to SEQ ID NO.: 3. Also comprised is the sequence of SEQ ID NO.: 4. Furthermore comprised is the sequence of SEQ ID NO.: 5 or a sequence which is at least 90% identical to the sequence of SEQ ID

20 NO.: 5. Preferably, said sequence is at least 95% identical to SEQ ID NO.: 5. More preferably, said sequence is at least 98% or 99% identical to SEQ ID NO.: 5. Also comprised is the sequence of SEQ ID NO.: 6 or a sequence which is at least 90% identical to the sequence of SEQ ID NO.: 6.

25 Preferably, said sequence is at least 95% identical to SEQ ID NO.: 6. More preferably, said sequence is at least 98% or 99% identical to SEQ ID NO.: 6. Likewise comprised is the sequence of SEQ ID NO.: 7 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 7.

30 Preferably, said sequence is at least 75% identical to SEQ ID NO.: 7. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 7. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 7. Also

35 comprised is the sequence of SEQ ID NO.: 8 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 8. Preferably, said sequence is at least 75% identical

to SEQ ID NO.: 8. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 8. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 8.

5

Additionally comprised is the sequence of SEQ ID NO.: 9 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 9. Preferably, said sequence is at least 75% identical to SEQ ID NO.: 9. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 9. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 9. Likewise comprised is the sequence of SEQ ID NO.: 10 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 10. Preferably, said sequence is at least 75% identical to SEQ ID NO.: 10. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 10. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 10. Also comprised is the sequence of SEQ ID NO.: 11 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 11. Preferably, said sequence is at least 75% identical to sequence SEQ ID NO.: 11. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 11. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 11. Furthermore comprised is the sequence of SEQ ID NO.: 12 or a sequence which is at least 60% identical to the sequence of SEQ ID NO.: 12. Preferably, said sequence is at least 65% identical to SEQ ID NO.: 12. More preferably, said sequence is at least 70% identical to the sequence of SEQ ID NO.: 12. Even more preferably, said sequence is at least 75% or 80% identical to SEQ ID NO.: 12. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 12. Likewise comprised is the sequence of SEQ ID NO.: 13 or a sequence which is at least

60% identical to the sequence of SEQ ID NO.: 13.  
Preferably, said sequence is at least 65% identical to SEQ ID NO.: 13. More preferably, said sequence is at least 70% identical to the sequence of SEQ ID NO.: 13. Even more  
5 preferably, said sequence is at least 75% or 80% identical to SEQ ID NO.: 13. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 13. Also comprised is the sequence of SEQ ID NO.: 14 or a sequence which is at least 75%  
10 identical to the sequence of SEQ ID NO.: 14. Preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 14. More preferably, said sequence is at least 85% identical to the sequence of SEQ ID NO.: 14. Even more preferably, said sequence is at least 90%, 95%, 98% or 99%  
15 identical to the sequence of SEQ ID NO.: 14. Additionally comprised is the sequence of SEQ ID NO.: 15 or a sequence which is at least 95% identical to the sequence of SEQ ID NO.: 15. Preferably, said sequence is at least 98% or 99% identical to the sequence of SEQ ID NO.: 15. Also comprised  
20 is the sequence of SEQ ID NO.: 16 or a sequence which is at least 95% identical to the sequence of SEQ ID NO.: 16. Preferably, said sequence is at least 98% or 99% identical to the sequence of SEQ ID NO.: 16. Furthermore comprised is the sequence of SEQ ID NO.: 17 or a sequence which is at  
25 least 75% identical to the sequence of SEQ ID NO.: 17. Preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 17. More preferably, said sequence is at least 85% identical to the sequence of SEQ ID NO.: 17. Even more preferably, said sequence is at least 90%,  
30 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 17. Likewise comprised is the sequence of SEQ ID NO.: 18 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 18. Preferably, said sequence is at least 75% identical to SEQ ID NO.: 18. More preferably, said  
35 sequence is at least 80% identical to the sequence of SEQ ID NO.: 18. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ

ID NO.: 18. Also comprised is the sequence of SEQ ID NO.: 19 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 19. Preferably, said sequence is at least 75% identical to SEQ ID NO.: 19. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 19. Even more preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 19. Additionally comprised is the sequence of SEQ ID NO.: 20 or a sequence which is at least 60% identical to the sequence of SEQ ID NO.: 20. Preferably, said sequence is at least 65% identical to SEQ ID NO.: 20. More preferably, said sequence is at least 70% identical to the sequence of SEQ ID NO.: 20. Even more preferably, said sequence is at least 75% or 80% identical to SEQ ID NO.: 20. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 20. Also comprised is the sequence of SEQ ID NO.: 21 or a sequence which is at least 90% identical to the sequence of SEQ ID NO.: 21. Preferably, said sequence is at least 95% identical to the sequence of SEQ ID NO.: 21. More preferably, said sequence is at least 98% or 99% identical to the sequence of SEQ ID NO.: 21. Furthermore comprised is the sequence of SEQ ID NO.: 22 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 22. Preferably, said sequence is at least 75% identical to SEQ ID NO.: 22. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 22. Particularly preferably said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 22. Likewise comprised is the sequence of SEQ ID NO.: 23 or a sequence which is at least 55% identical to the sequence of SEQ ID NO.: 23. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 23. More preferably, said sequence is at least 65% identical to the sequence of SEQ ID NO.: 23. Even more preferably, said sequence is at least 70% or 75% identical to SEQ ID NO.: 23. Particularly preferably, said sequence is at least 80%, 85%, 90%, 95%, 98% or 99%

identical to the sequence of SEQ ID NO.: 23. Also comprised is the sequence of SEQ ID NO.: 24 or a sequence which is at least 65% identical to the sequence of SEQ ID NO.: 24. Preferably, said sequence is at least 70% identical to SEQ ID NO.: 24. More preferably, said sequence is at least 75% identical to the sequence of SEQ ID NO.: 24. Even more preferably, said sequence is at least 80% or 85% identical to SEQ ID NO.: 24. Particularly preferably, said sequence is at least 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 24. Furthermore comprised is the sequence of SEQ ID NO.: 25 or a sequence which is at least 55% identical to the sequence of SEQ ID NO.: 25. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 25. More preferably, said sequence is at least 65% identical to the sequence of SEQ ID NO.: 25. Even more preferably, said sequence is at least 70% or 75% identical to SEQ ID NO.: 25. Particularly preferably, said sequence is at least 80%, 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 25. Also comprised is the sequence of SEQ ID NO.: 26 or a sequence which is at least 55% identical to the sequence of SEQ ID NO.: 26. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 26. More preferably, said sequence is at least 65% identical to the sequence of SEQ ID NO.: 26. Even more preferably, said sequence is at least 70% or 75% identical to SEQ ID NO.: 26. Particularly preferably, said sequence is at least 80%, 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 26. Likewise comprised is the sequence of SEQ ID NO.: 27 or a sequence which is at least 55% identical to the sequence of SEQ ID NO.: 27. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 27. More preferably, said sequence is at least 65% identical to the sequence of SEQ ID NO.: 27. Even more preferably, said sequence is at least 70% or 75% identical to SEQ ID NO.: 27. Particularly preferably, said sequence is at least 80%, 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 27. Furthermore comprised is the sequence of SEQ ID NO.: 28 or a sequence

which is at least 75% identical to the sequence of SEQ ID NO.: 28. Preferably, said sequence is at least 80% identical to SEQ ID NO.: 28. More preferably, said sequence is at least 85% identical to the sequence of SEQ ID NO.:

5 28. Even more preferably, said sequence is at least 90%, 95%, 98% or 99% identical to SEQ ID NO.: 28. Likewise comprised is the sequence of SEQ ID NO.: 29 or a sequence which is at least 70% identical to the sequence of SEQ ID NO.: 29. Preferably, said sequence is at least 75%

10 identical to SEQ ID NO.: 29. More preferably, said sequence is at least 80% identical to the sequence of SEQ ID NO.: 29. Even more preferably, said sequence is at least 85% or 90% identical to SEQ ID NO.: 29. Particularly preferably, said sequence is at least 95%, 98% or 99% identical to the

15 sequence of SEQ ID NO.: 29. Also comprised is the sequence of SEQ ID NO.: 30 or a sequence which is at least 60% identical to the sequence of SEQ ID NO.: 30. Preferably, said sequence is at least 65% identical to SEQ ID NO.: 30. More preferably, said sequence is at least 70% identical to

20 the sequence of SEQ ID NO.: 30. Even more preferably, said sequence is at least 75% or 80% identical to SEQ ID NO.: 30. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 30. Furthermore comprised is the sequence of SEQ ID

25 NO.: 31 or a sequence which is at least 55% identical to the sequence of SEQ ID NO.: 31. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 31. More preferably, said sequence is at least 65% or 70% identical to the sequence of SEQ ID NO.: 31. Even more preferably,

30 said sequence is at least 75% or 80% identical to SEQ ID NO.: 31. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 31. Likewise comprised is the sequence of SEQ ID NO.: 32 or a sequence which is at least 55% identical to

35 the sequence of SEQ ID NO.: 32. Preferably, said sequence is at least 60% identical to SEQ ID NO.: 32. More preferably, said sequence is at least 65% or 70% identical



to the sequence of SEQ ID NO.: 32. Even more preferably, said sequence is at least 75% or 80% identical to SEQ ID NO.: 32. Particularly preferably, said sequence is at least 85%, 90%, 95%, 98% or 99% identical to the sequence of SEQ ID NO.: 32. Furthermore comprised is the sequence of SEQ ID NO.: 33 or the sequence of SEQ ID NO.: 34. The amino acid sequences mentioned and characterized by SEQ IDs are preferably encoded by the DNA sequences referred to as SEQ ID numbers 35 to 68. Preference is further given to polypeptides which correspond to the naturally occurring enzymes over their full length. In another preferred embodiment, the polypeptides of the invention additionally comprise at least one heterologous amino acid section which characterizes said polypeptides as fusion proteins.

Possible examples of heterologous components of the fusion protein of the invention are Tags (e.g. His-Tag or Flag-Tag) which may be used for purification of the fusion proteins of the invention. In other embodiments, the heterologous components may have their own enzymic activity. In such a case, the two enzymic components are preferably linked by a linker such as a flexible glycine or glycine-serine linker of 6-10 amino acids in length, in order to ensure the functionality of said components. The term "heterologous", as used herein, may mean, firstly, that the components of the fusion protein do not naturally occur covalently linked together and, secondly, that the components originate from different species. Fusion proteins are usually prepared using recombinant DNA technology (see Sambrook et al., loc. cit.).

According to the invention, the term "polypeptide which has the biological activity of an NAD- or NADP-dependent alcohol dehydrogenase" refers to a group of enzymes which catalyze the conversion of alcohols to aldehydes or ketones or the corresponding reverse reaction, i.e. the conversion of aldehydes to primary alcohols or ketones to secondary alcohols. The first-mentioned reaction corresponds in this

connection to an oxidative process, with the secondly mentioned type of reaction being a reductive process. The EC number of alcohol dehydrogenases (ADHs) is EC 1.1.1.1. The scope of protection of the invention comprises, in addition to the naturally occurring enzymes isolated in the course of the present invention, also those polypeptides which have the aforementioned identity values at the amino acid level compared to the polypeptides isolated from natural sources and which may likewise originate from natural sources. On the other hand, they may be modified by recombinant DNA technology in such a way that the enzymic activity is retained or essentially retained, as will be anticipated by the skilled worker (cf., for example, Sambrook et al, "Molecular Cloning, A Laboratory Handbook", 2nd edition 1989, CSH Press, Cold Spring Harbor, Ausubel et al. "Current Protocols in Molecular Biology", John Wiley & Sons, NY 2001). Thus, it is possible for amino acids which are not located at the active site and whose replacement with an amino acid "of the same kind" is not expected at first sight to result in a substantially altered three-dimensional structure to be replaced with an amino acid "of the same kind". For example, particular amino acids with nonpolar side chains (amino acids of the same kind), may be expected to be able to be substituted, for example valine for alanine, without this having a (substantial) influence on the biological function of the enzyme, on the enzymic activity in accordance with the invention. On the basis of his specialist knowledge, the skilled worker may draw corresponding conclusions also for the substitution of other types of amino acids (for example the replacement of basic amino acids with other basic amino acids or of amino acids with uncharged polar side chains with other amino acids from this group).

The percentage of identity to the amino acid sequences of the polypeptides isolated from natural sources, which are described in this description by SEQ ID numbers, may be

readily determined by the skilled worker using processes known in the prior art. A suitable program which may be used according to the invention is BLASTP (Altschul et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-3402.).

The invention also relates to a nucleic acid molecule which encodes the polypeptide of the invention.

10

The nucleic acid molecule of the invention may be a DNA molecule or an RNA molecule. Preference is given to the nucleic acid molecule being a cDNA molecule or an mRNA molecule. According to the invention, said DNA molecule may furthermore be a genomic DNA molecule. The invention also comprises embodiments in which said DNA molecule is a PNA molecule or another derivative of a DNA molecule. According to the invention, preference is given to DNA sequences comprising the DNA sequences according to SEQ ID numbers 35 to 68.

20

In order to achieve the object on which the invention is based, the following approach was pursued. Firstly, based on an extensive proprietary strain collection, prioritized strains were grown on plates and, after viability and purity controls had been carried out, also in liquid culture. The genomic DNA of these organisms was isolated from the harvested cell pellets. Based on the genomic DNA prepared, selected isolates were genetically screened for alcohol dehydrogenase genes via PCR typing by means of primers of the invention. In this context, even the amino acid sequence similarity due to homology of already known alcohol dehydrogenases did not readily allow oligonucleotide primers to be derived with the aid of which previously unidentified alcohol dehydrogenase genes may readily be amplified successfully. Initially, this approach was based on the hypothesis of particular sequence motifs

30

35

conserved in the previously known alcohol dehydrogenase genes also being present in the desired novel alcohol dehydrogenase genes. However, the sequence motifs conserved in the previously known alcohol dehydrogenase genes are unsuitable for deriving degenerated primers by processes known to the skilled worker (Kwok et al., 1995. Design and use of mismatched and degenerate primers. In: PCR Primer, A laboratory Manual, Dieffenbach CW & Dveksler GS (Editors), Cold Spring Harbor Laboratory Press, pp143-155; Compton T. 1990. Degenerate Primers for DNA Amplification. In: PCR Protocols, A Guide to Methods and Applications. Innis et al. (Editors) Academic Press, San Diego, pp39-34). The NAD- or NADP-dependent alcohol dehydrogenases are classified as long-chain, medium-chain and short-chain ADHs. They are divided into these three groups especially based on their metal dependence and the size of subunits. The short-chain ADHs do not require any metal ions and their subunits consist of approximately 250 amino acids. In contrast, the medium-chain and long-chain ADHs are dependent on metal ions. The medium-chain ones whose typical subunits consist of approx. 350 amino acids require zinc ions. The long-chain ADHs whose subunits are composed of approx. 385 amino acids require iron ions (Hummel, W. 1997. New alcohol dehydrogenases for the synthesis of chiral compounds. 58:145-84). The sequence heterogeneity, not only within all of the previously known NAD- or NADP-dependent ADHs but also within the three ADH groups briefly described above, is extremely high. Therefore, it was not that easy to construct primers with the aid of which it is possible firstly to amplify specifically ADH sequences and secondly also to capture a diversity of novel ADHs necessary in order to achieve the object. Thus, despite the sequence homologies expected on the basis of the used, no long-chain ADHs whatsoever were isolated in the bacteria studied. In this connection, it was intended to test the quality of the constructed primers first with genomic DNA of model organisms whose alcohol dehydrogenase genes are

known or with DNA pools consisting of DNA from various microorganisms. This involved cloning, sequencing and subsequently analyzing PCR products. After this establishing phase, selected isolates were subjected to PCR  
5 typing on the basis of the prepared genomic DNA of the microorganisms to be screened, as described above. The results obtained from the experiments (regarding sequence identity and specific activity) were incorporated into prioritizing the potential Hit organisms whose novel  
10 alcohol dehydrogenase genes are being isolated. Despite the unexpected, disappointing and demotivating results in the course of the attempted isolation of nucleic acids supposed to encode long-chain ADHs, a number of nucleic acids encoding short-chain enzymes and medium-chain enzyme chains  
15 were isolated according to the invention. Some of these enzymes and enzyme chains had surprisingly low sequence identities (<50%) to the known enzymes of this class.

The invention furthermore relates to a nucleic acid  
20 molecule which is complementary to the nucleic acid molecule of the invention.

According to the invention, the term "complementary" means a complementarity which extends across the entire region of  
25 the nucleic acid molecule of the invention without gaps. In other words, preference is given according to the invention to said complementarity extending 100% across the entire region of the sequence of the invention, i.e. from the 5' end shown to the 3' end shown. In further preferred  
30 embodiments, said complementarity extends across a region of at least 19, preferably at least 21, contiguous nucleotides which preferably do not code for the active site of enzymic activity.

35 In addition, the invention relates to a vector which comprises the nucleic acid molecule of the invention.

The vectors of the invention preferably contain the nucleic acids of the invention operatively linked to an expression control sequence so as for said nucleic acids to be able to be transcribed and, where appropriate, translated in a suitable host cell. Expression control sequences usually comprise a promoter and, where appropriate, further regulatory sequences such as operators or enhancers. Furthermore, translation initiation sequences may also be present. Suitable expression control sequences for prokaryotic or eukaryotic host cells are known to the skilled worker (see, for example, Sambrook et al., loc. cit.). The recombinant vector of the invention may furthermore also contain usual elements such as an origin of replication and a selection marker gene. Examples of suitable recombinant vectors are plasmids, cosmids, phages or viruses (see, for example, Sambrook et al., supra). Starting materials for preparing the recombinant vectors of the invention are commercially available (for example from Stratagene, InVitroGen or Promega).

Any plasmids or vectors available to the skilled worker for this purpose are suitable in principle. Plasmids and vectors of this kind may be found, for example, in Studier and coworkers (Studier, W. F.; Rosenberg A. H.; Dunn J. J.; Dubendroff J. W.; (1990), Use of the T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185, 61-89) or in the brochures from Novagen, Promega, New England Biolabs, Clontech or Gibco BRL. Further preferred plasmids and vectors may be found in: Glover, D. M. (1985), DNA cloning: a practical approach, Vol. I-III, IRL Press Ltd., Oxford; Rodriguez, R.L. and Denhardt, D. T (eds) (1988), Vectors: a survey of molecular cloning vectors and their uses, 179-204, Butterworth, Stoneham; Goeddel, D. V. (1990), Systems for heterologous gene expression, Methods Enzymol. 185, 3-7; Sambrook, J.; Fritsch, E. F. and Maniatis, T. (1989), Molecular cloning: a laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, New

York. Plasmids which may be used for cloning the gene construct having the nucleic acid of the invention into the host organism in a very preferred manner are derivatives of: pUC18 and pUC19 (Roche Biochemicals), pKK-177-3H (Roche Biochemicals), pBTac2 (Roche Biochemicals), pKK223-3 (Amersham Pharmacia Biotech), pKK-233-3 (Stratagene) or pET (Novagen). Other preferred plasmids are pBR322 (DSM3879), pACYC184 (DSM4439) and pSC101 (DSM6202), which may be obtained from the DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Brunswick, Germany. Examples of preferred promoters are T7, lac, tac, trp, rha and ara.

The invention also relates to a nonhuman host which comprises the polypeptide of the invention or the nucleic acid molecule of the invention or the vector of the invention.

The nonhuman host may be a cell or a multi- to polycellular organism. Suitable polycellular organisms include model systems familiar in molecular biology, such as *Drosophila melanogaster*, zebra fish or *C. elegans*.

In a preferred embodiment, the host is a cell.

In this preferred embodiment, the host of the invention is a recombinant cell which has been transformed or transfected with a nucleic acid of the invention or a vector of the invention (according to the present invention, the terms "transformation" and "transfection" are used synonymously). Transformation or transfection may be carried out by known methods, for example calcium phosphate coprecipitation, lipofection, electroporation, particle bombardment or viral infection. The cell of the invention may contain the recombinant nucleic acid in an extrachromosomal or a chromosomally integrated form. In other words, the transfection/transformation may be stable or transient.

The recombinant cell preferably is of prokaryotic origin. Suitable host cells include cells of unicellular microorganisms, such as bacteria cells. A particularly suitable bacterial host system is *E. coli*. The cytoplasm of *E. coli* contains the cofactors required for the enzymic activity of the polypeptide of the invention. These are, in particular, NADH, NADPH, NAD<sup>+</sup> and NADP<sup>+</sup>. Very particular preference is given to: *E. coli* XL1 Blue, W3110, DSM14459 (PCT/US00/08159), NM 522, JM101, JM109, JM105, RR1, DH5, TOP 10- or HB101. It is also possible to use for expression of the nucleic acids of the invention bacteria of the genera/species *Lactobacillus*, *Bacillus*, *Rhodococcus*, *Campylobacter*, *Caulobacter*, *Mycobacterium*, *Streptomyces*, *Neisseria*, *Ralstoni*, *Pseudomonas*, and *Agrobacterium*. Appropriate strains are available in the prior art and may, at least partially, be obtained via the international deposition sites such as ATCC or DMSZ. Transfection protocols and transformation protocols are known to the skilled worker. (Chan and Cohen. 1979. High Frequency Transformation of *Bacillus subtilis* Protoplasts by Plasmid DNA. *Mol Gen Genet.* 168(1):111-5; Kieser et al.. 2000. *Practical Streptomyces Genetics*. The John Innes Foundation Norwich.; Sambrook et al.. 1989. *Molecular Cloning. A Laboratory Manual*. In: second ed.. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. NY.; Irani and Rowe. 1997. Enhancement of transformation in *Pseudomonas aeruginosa* PAO1 by Mg<sup>2+</sup> and heat. *Biotechniques* 22: 54-56). Additional examples of host organisms which may be used are also yeasts such as *Hansenula polymorpha*, *Pichia sp.*, *Saccharomyces cerevisiae*. As an alternative to this, the cell may be of eukaryotic origin. Suitable eukaryotic cells include CHO cells, HeLa cells and others. Many of these cells are obtainable via deposition sites such as ATCC or DMSZ.

In a further preferred embodiment, the host is a transgenic nonhuman animal.



Transgenic nonhuman animals may be produced by processes known in the prior art.

The transgenic nonhuman animal of the invention may  
5 preferably have various genetic constitutions. It may (i) constitutively or inducibly overexpress the gene of a nucleic acid of the invention, (ii) contain the endogenous gene of a nucleic acid of the invention in an inactive form, (iii) contain the endogenous gene of a nucleic acid  
10 of the invention completely or partially replaced with a mutated gene of a nucleic acid of the invention, (iv) have conditional and tissue-specific overexpression or underexpression of the gene of a nucleic acid of the invention or (v) have a conditional and tissue-specific  
15 knock-out of the gene of a nucleic acid of the invention. Preferably, the transgenic animal additionally contains an exogenous gene of a nucleic acid of the invention under the control of a promoter allowing overexpression. Alternatively, the endogenous gene of a nucleic acid of the  
20 invention may be overexpressed by activating or/and replacing the intrinsic promoter. Preferably, the endogenous promoter of the gene of a nucleic acid of the invention has a genetic modification which results in increased expression of the gene. Said genetic modification  
25 of the endogenous promoter here comprises both a mutation of individual bases and deletion and insertion mutations.

In a particularly preferred embodiment of the host of the invention, the latter is a transgenic rodent, preferably a  
30 transgenic mouse, a transgenic rabbit, a transgenic rat, or is a transgenic sheep, a transgenic cow, a transgenic goat or a transgenic pig.

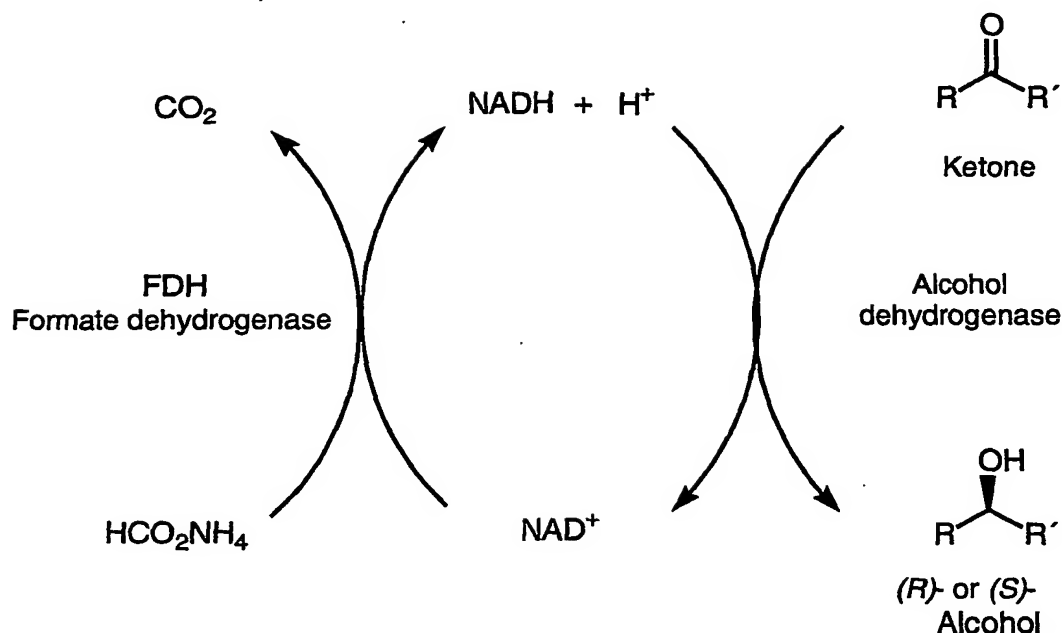
Mice have numerous advantages over other animals. They can  
35 be kept easily and their physiology is regarded as a model system for that of humans. The production of such gene-manipulated animals is sufficiently known to the skilled

worker and carried out using common processes (see, for example, Hogan, B., Beddington, R., Costantini, F. and Lacy, E. (1994), *Manipulating the Mouse-Embryo; A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; WO91/08216).

Alternatively or additionally, it is also possible to employ cell culture systems, in particular human cell culture systems, for the applications described for the nonhuman transgenic animal of the invention.

10 Examples of cofactors of alcohol dehydrogenases, which are used - as already mentioned, depending on the particular alcohol dehydrogenase - are NADH and NADPH and their oxidized forms,  $\text{NAD}^+$  and  $\text{NADP}^+$ , respectively.

15 The cofactors may be regenerated, in principle, either in an enzyme-coupled manner using a second enzyme, for example a formate dehydrogenase or glucose dehydrogenase, or in a substrate-coupled manner using any of the alcohols accepted as substrate by the alcohol dehydrogenase  
20 employed, for example - *iso*-propanol - if accepted as substrate. The diagram below indicates by way of example the concept of the alcohol dehydrogenase-catalyzed reduction of a ketone with enzyme-coupled cofactor regeneration using a formate dehydrogenase.



Alcohol dehydrogenases are used, for example, for preparing enantiomerically enriched, preferably enantiomerically pure, secondary alcohols, starting from prochiral ketones. In this connection, both (R)- and (S)-specific alcohol dehydrogenases are known which accordingly result in the formation of the particular enantiomeric (R) and, respectively, (S) forms of alcohols.

10

Correspondingly, preference is also given according to the invention to a host which has a further dehydrogenase suitable for cofactor regeneration or a nucleic acid molecule encoding said dehydrogenase.

15

In this connection, the host may contain naturally said further dehydrogenase or may have been transfected with a recombinant nucleic acid which encodes said dehydrogenase and with which it is possible to express said dehydrogenase in said host. This embodiment also requires the host having the cofactors necessary for the function of said further dehydrogenase or said cofactors being delivered to said host in a suitable manner.

20

Particular preference is given in this connection to a host in which the dehydrogenase suitable for cofactor regeneration is a formate dehydrogenase or a glucose dehydrogenase. Particular preference is given to a *Candida boidinii* formate dehydrogenase. Particular preference is also given to the cofactor-regenerating dehydrogenase being a *Bacillus subtilis* glucose dehydrogenase. Genetically modified mutants of said cofactor-regenerating dehydrogenases, which retain said enzymic function, are likewise preferred according to the invention.

In a further embodiment, the invention relates to a reaction system which comprises an organic compound which is a substrate of a dehydrogenase, furthermore the polypeptide of the invention, the vector of the invention or the host of the invention and, where appropriate, a cofactor for the polypeptide of the invention. (The addition of cofactor is required in those cases in which the cofactor is not already present in the system, see also hereinbelow). In one case, the reaction system of the invention may be a bacterial cell which corresponds to the host of the invention and which has the polypeptide of the invention and also the necessary cofactors in the cytoplasm. In the case of the cofactor(s) already being present naturally in the system/host, said cofactors need no longer be delivered separately. Suitably, the host is one which has a further dehydrogenase suitable for cofactor regeneration or a nucleic acid molecule encoding said dehydrogenase and also the cofactors required therefor. If a substrate for a desired product is supplied to said reaction system or if said substrate is metabolized in the reaction system itself, then the desired product may readily be isolated from the reaction system, if said reaction system is maintained under suitable conditions. Suitable conditions include carrying out the reaction at temperatures of from 10 to 80°C, preferably from 20 to 60°C, and very preferably from 20 to 40°C. Preference is

also given to the substrate concentration being from 100 to 2000 mM, preferably from 200 to 800 mM. In a preferred form, the desired reaction is carried out so as to achieve conversions of >80%, in particular >90%, within a reaction time of <20 hours, in particular a reaction time of <10 hours and very preferably a reaction time of <5 hours. In another embodiment, the reaction system may be an in vitro system for converting a suitable substrate to obtain the desired product. For example, the polypeptide of the invention may be contacted with the cofactors mentioned and the substrate and, where appropriate (i.e. if necessary), with a further dehydrogenase suitable for cofactor regeneration (and, where appropriate, cofactors required therefor, in particular NADH and/or NADPH their oxidized forms) under suitable conditions, as set out above, for example, and over a sufficient period of time, so that the desired product may be generated. In this in vitro variant with utilization of isolated enzymes (in purified form or as crude extract) and addition of cofactors, these cofactor additions should, in accordance with an economical process control, be <0.01 equivalents (based on the amount of substrate employed), preferably <0.001 equivalents and very preferably <0.0005 equivalents.

The "reaction system" may moreover also be a transgenic nonhuman animal to which a suitable substrate and, where appropriate, cofactors or/and said further dehydrogenase is fed or administered and which is capable of converting said substrate in suitable tissues. In another embodiment, the reaction system may also be a cellular membrane system in which the enzyme, the enzymes and, where appropriate, the cofactors are anchored.

Further preference is given according to the invention to a reaction system in which the organic compound which is a substrate of a dehydrogenase is a carbonyl compound.

Particular preference is given to a reaction system in which the carbonyl compound is an aldehyde or a ketone.

5 This embodiment of the invention permits the preparation, particularly preferred according to the invention, of technical grade alcohols which may be used, for example, as intermediates for the preparation of active compounds usable in medicaments.

10 Particular preference is given according to the invention to the ketone being an asymmetrically substituted ketone.

This embodiment of the invention is particularly preferred because the products generated in a corresponding reduction  
15 have a center of chirality and may be obtained with high enantioselectivity. In general, the desired chiral secondary alcohols are obtained in an enantiomerically pure form with an enantiomeric excess of >99%.

20 In another preferred embodiment of the reaction system of the invention, the organic compound which is a substrate of a dehydrogenase is an alcohol. This variant is preferably suitable for preparing commercially important carbonyl compounds, for example ketones relevant in the field of  
25 aroma chemicals. In addition, oxidation may also be utilized for the formation of enantiomerically pure, secondary alcohols by starting from a racemic alcohol as substrate and converting the undesired enantiomer into the ketone compound by enantioselective oxidation. The  
30 remaining, desired enantiomer may then be isolated accordingly.

The alcohol is preferably a primary alcohol or a chiral secondary alcohol. In the first case, the product generated is an aldehyde, whereas in the second case the  
35 corresponding ketones are formed.

Preference is also given according to the invention to the cofactor in the reaction system of the invention being NADH, NADPH, NAD<sup>+</sup> or NADP<sup>+</sup>.

- 5 The invention also relates to a process for preparing the polypeptide of the invention or a polypeptide encoded by the nucleic acid molecule of the invention, which process comprises growing the host of the invention and isolating said polypeptide.

10

- The polypeptide may be purified, for example, by conventional processes, for example by disrupting appropriate cells, for example by means of a "French press", by ion exchange, size selection or affinity chromatography etc. (Coligan et al.. Current Protocols in Protein Science, John Wiley & Sons, Inc.). As an alternative to this, the polypeptide of the invention, when linked to a leader peptide, may be exported out of the cells and purified from the culture supernatant. This embodiment requires the polypeptide of the invention, which does not naturally contain a leader peptide, to be genetically modified. This embodiment has the advantage of a simpler purification of the polypeptide of the invention from the culture supernatant. The best procedures and suitable leader peptides may be readily determined by the skilled worker.
- 15  
20  
25

- In a further preferred embodiment of the process of the invention, the polypeptide is isolated from a body fluid or tissue sample of the nonhuman transgenic animal. In this embodiment too, the polypeptide of the invention preferably contains a leader peptide.
- 30

- In a further preferred embodiment of the process of the invention and, in particular, if the nonhuman transgenic animal is a mammal, for example a cow, a goat or a sheep, the body fluid is milk or serum.
- 35

In another embodiment, the invention relates to a process for preparing an organic compound which is a product of a dehydrogenase, which process comprises reacting an organic compound which is a substrate of a dehydrogenase with the polypeptide of the invention, the host of the invention or by means of the reaction system of the invention.

The various embodiments of the invention which are to be used in the process of the invention differ in principle in that further components such as cofactors etc. (cf. supra) have to be added to the polypeptide, if the latter is used in a cell-free in vitro system. When using the reaction system of the invention, the necessary components, with the possible exception of the substrate, are preferably and advantageously already present in the system, and a separate addition is thus not needed here.

Preference is given according to the invention to a process which further comprises the step of isolating the product of the reaction. Suitable processes for isolation/purification have been set forth above.

In a particularly preferred embodiment of the process of the invention, the latter further comprises processing the product to give a medicament. A number of descriptions of utilizing enantiomerically pure alcohols as intermediates for preparing pharmaceutical active compounds are given in the literature. An overview in this respect is contained, inter alia, in: A. Kleemann, J. Engels, B. Kutscher, D. Reichert, Pharmaceutical Substances: Syntheses, Patents, Applications, 4th edition, Thieme-Verlag, Stuttgart, 2001.

In another, particularly preferred embodiment of the process of the invention, the latter further comprises the step of processing the product to give a secondary product. In this connection, derivatization may take place both by way of modification of the alcohol group, for example by



esterification and subsequent secondary reactions, and by way of modifications of the particular substituents.

- Particular preference is given here to the process of the invention further comprising the step of formulating the secondary product with a pharmaceutically compatible carrier or excipient or diluent in the preparation of a medicament.
- Examples of suitable pharmaceutically compatible carriers and/or diluents are known to the skilled worker and comprise, for example, phosphate-buffered saline solutions, water, emulsions such as, for example, oil/water emulsions, various types of wetting agents or detergents, sterile solutions, etc. Medicaments comprising such carriers may be formulated by means of known conventional methods. Said medicaments may be administered in a suitable dose to an individual. The administration may be carried out orally or parenterally, for example intravenously, intraperitoneally, subcutaneously, intramuscularly, locally, intranasally, intrabronchially, orally or intradermally, or via a catheter at a site in an artery. Preparations for parenteral administration comprise sterile aqueous or nonaqueous solutions, suspensions and emulsions. Examples of nonaqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as, for example, olive oil, and organic ester compounds such as, for example, ethyl oleate, which are suitable for injections. Aqueous carriers comprise water, alcohol/water-based solutions, emulsions, suspensions, salt solutions and buffered media. Parenteral carriers comprise sodium chloride solutions, Ringer dextrose, dextrose and sodium chloride, Ringer lactate and bound oils. Examples of intravenous carriers include liquid, nutrient and electrolyte supplements (such as, for example, those based on Ringer dextrose). The medicament may moreover comprise preservatives and other additives such as, for example, antimicrobial compounds, antioxidants,

complexing agents and inert gases. Depending on the intended specific usage, other active compounds such as, for example, interleukins, growth factors, differentiation factors, interferons, chemotactic proteins or an unspecific immunomodulating agent, may also be included.

The type of dosage is determined by the attending physician according to the clinical factors. The skilled worker knows that the type of dosage depends on various factors such as, for example, body size or weight, body surface area, age, sex or general health of the patient, or else on the agent to be specially administered, duration and type of administration, and on other medicaments which may be administered in parallel. A typical dose may be, for example, in a range between 0.001 and 1000  $\mu\text{g}$ , with doses being conceivable below and above this exemplary range, especially when taking into account the abovementioned factors. If the composition of the invention is administered regularly, the unit dose per day should generally be in a range between 1  $\mu\text{g}$  and 10 mg. The active compounds in these preparations are usually present at a concentration of more than 10  $\mu\text{g}/\text{ml}$  of a physiological buffer. However, they may also be present in solid form at a concentration of from 0.1 to 99.5% by weight of the total mixture. In general, it has proven advantageous to administer the active compound(s) in total amounts of from about 0.001 to 100 mg/kg, preferably in total amounts of from about 0.01 to 10 mg/kg, of body weight per 24 hours, where appropriate as continuous infusion or in the form of a plurality of individual doses, in order to achieve the desired result. If the composition is administered intravenously, the unit dose per kilogram of body weight per day should be in a range between 1  $\mu\text{g}$  and 10 mg. The medicament may be administered topically, locally or systemically.

Finally, particular preference is given according to the invention to a process in which the product is an enantiomerically pure alcohol.

- 5 The invention also relates to a ligand which specifically binds the polypeptide of the invention, which ligand is neither a substrate of said polypeptide, nor a cofactor thereof, nor a product converted thereby.
- 10 The term "specifically binds" means according to the invention that the ligand does not or essentially does not cross react with other polypeptides, including those having a similar primary sequence or a similar three-dimensional structure. Cross reactivity may be determined by processes
- 15 known in the prior art (cf. Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988). To this end, it is possible to use, for example, competitive assays, for example turbidimetric tests, in which the ligand is incubated together with the labeled
- 20 polypeptide of the invention and a polypeptide competing therewith, it being possible for the latter to be used at different concentrations.

- In a preferred embodiment, the ligand of the invention is
- 25 an antibody or a fragment or derivative thereof, an aptamer, or a low-molecular weight substance.

- Antibody fragments comprise Fv, Fab and F(ab')<sub>2</sub> fragments. The derivatives include scFvs (Harlow and Lane, loc. cit.).
- 30 Antibodies may be of polyclonal or monoclonal origin. In a particularly preferred embodiment of the receptor of the invention, said receptor is a monoclonal antibody.

- According to the invention, "low molecular weight
- 35 substances" are naturally occurring or artificially produced molecules having a molecular weight of from about 250 to 1000 Da, preferably 300 to 750 Da, particularly

preferably 400 to 600 Da, or are modified molecules of said molecular weight, which have been derived from natural substances.

- 5 The claimed invention furthermore comprises a primer having a sequence depicted in Table 1.

In addition, the invention relates to a primer pair having sequences depicted in Table 1, with the first primer of  
10 said primer pair serving as a forward primer and the second primer of said primer pair serving as a reverse primer to amplify a DNA sequence.

The primer of the invention (in combination with a further  
15 suitable primer, preferably a further suitable primer listed in Table 1) and the primer pair of the invention may be used for amplification of the sequences of the invention, preferably by means of PCR or LCR. The primers and primer pairs, respectively, have been selected with  
20 great care from a multiplicity of potentially possible primers. Besides amplification of the nucleic acid sequences of the invention, they also allow amplification of sequences which encode enzymes of the prior art, and are thus versatile.

25

The invention further relates to a kit comprising

the polypeptide of the invention;

30 the nucleic acid molecule of the invention;

the vector of the invention;

the host of the invention;

35

the ligand of the invention;

the reaction system of the invention;

at least one primer of the invention; and/or

5 at least one primer pair of the invention.

The components of the kit of the invention may be packaged individually or partly together in suitable vessels. The components may be present in the kit of the invention, for example, in freeze-dried form or, for example, in solution,  
10 with suitable solvents including in particular aqueous solvents such as buffered solutions, for example phosphate-buffered solutions.

The kits of the invention may be used in many different  
15 ways. For example, they may serve to identify further alcohol dehydrogenases or nucleic acids encoding these, with preference being given to using the primers of the invention. In other embodiments, the kits of the invention may be used for industrial production of the enzyme of the  
20 invention or of the products converted by said enzyme. In these embodiments, preference would be given to using the host of the invention or the reaction system of the invention.

In the Figures:

25 Fig. depicts the prior art via the resolution of the racemate: at least 4-4 steps

Fig. depicts an overview of cluster 2  
(= primer group 2), based on 33 sequences

Fig. 3: depicts PCR typing with primer group 2,  
30 using various pools

The examples illustrate the invention. Example 1: Clustering of ADHs and primer design

35 Strains were prioritized from an extensive proprietary strain collection and grown in liquid culture after carrying out viability and purity checks. Harvested cell

pellets served as starting material for genetic screening. Primers for genetic screening for alcohol dehydrogenase genes were constructed and then tested on the basis of prepared genomic DNA of selected microbial isolates with the aid of PCR. NAD- or NADP-dependent alcohol dehydrogenases are classified as long-chain, medium-chain and short-chain ADHs. Since sequence heterogeneity within these groups is substantial, said groups were grouped in clusters based on sequence analyses. The long-chain ADHs were divided into three clusters, the medium-chain ADHs in 4 clusters and the short-chain ADHs in three clusters. Subsequently, in each case four degenerated primer sets per cluster were constructed which differ in the utilization of specific codons (codon usage) but which are directed against the same sequence motifs of the clusters.

Example 2:

#### Genetic screening for long-chain alcohol dehydrogenases

The long-chain ADHs were divided on the basis of sequence analyses into three clusters and the primers were constructed and tested with an analogous procedure. However, despite expectations to the contrary, no PCR tags assignable to this group were amplified.

Example 3:

#### Genetic screening for medium-chain alcohol dehydrogenases

Primers directed against medium-chain ADHs were designed as follows: the medium-chain ADHs were divided based on sequence analysis into four clusters. Subsequently, in each case four degenerated primer sets were constructed which differ by the selected codon usage. This will be illustrated graphically and by way of example in Figure 2 on the basis of the group of organisms for determining the primer group 2. The primer sets were selected on the basis of conserved regions in 33 different alcohol dehydrogenase sequences.

These primer groups, for example primer group 2, were subsequently used for investigating various pools (containing genomic DNA from microorganisms). Using this primer set, it was possible to amplify, clone and sequence novel partial medium-chain ADH sequences. The corresponding result of this PCR typing is depicted below in Figure 3. As documented, inter alia, by lanes 1, 2 and 10 in Figure 3, in each case here gene sequences were found which indicate an alcohol dehydrogenase activity, owing to the gene sequence corresponding to known genes of ADH enzymes. Overall, further gene sequences with potential alcohol dehydrogenase activity were identified. The identity of the sequence tags found with already known ADHs was between 51 - 99%.

Example 4:

Genetic screening for special medium-chain alcohol dehydrogenases, with analogy to alcohol dehydrogenases from *Rhodococcus* strains

Owing to the interesting properties of the known *Rhodococcus erythropolis* (S)-alcohol dehydrogenase (S-Re-ADH; this enzyme is characterized by stereoselective conversion of a broad spectrum of ketones and ketoesters to the corresponding hydroxy compounds) which likewise belongs to the medium-chain alcohol dehydrogenases, and also of other alcohol dehydrogenases obtained from *Rhodococcus* strains, the question as to whether it is possible to identify novel ADH sequences exhibiting a relatively high similarity to this sequence with the aid of genetic screening, was looked into. The driving force here is the assumption that novel ADHs whose sequences share a high identity with the S-Re-ADH sequence could likewise possess interesting properties. For example, such novel ADHs could possess on the one hand the proven properties of S-Re-ADH but, on the other hand, for example, could have a modified substrate spectrum or increased expression performance. In order to answer the above question, comparative sequence analyses with the amino acid sequence of this S-Re-ADH were

carried out first. These analyses revealed that S-Re-ADH is a representative of cluster 1 of the medium-chain ADHs. Furthermore, a group consisting of 5 protein sequences, which includes S-Re-ADH, was found within this cluster.

- 5 Starting from these 5 sequences, degenerated primers were constructed and assayed, taking into account the codon usage, according to the procedure described above.

- 10 In order to reduce the number of PCRs to be carried out, pools consisting of 24 bacterial isolates were established and DNA was isolated. This DNA was used as template. Numerous sequence tags were amplified and sequenced. Analysis of the sequence tags translated into amino acid sequences revealed identities to the S-Re-ADH sequence of  
15 from about to Two full-length genes were isolated which are represented by one sequence tag and which exhibit 98% identity to S-Re-ADH at the amino acid sequence level. The novel ADHs are derived from the organism *Arthrobacter paraffineus* ATCC21317. The homology at the DNA level is  
20 94%. Said full-length genes were isolated with the aid of a sequence homology approach.

#### Example 5:

- 25 Genetic screening for short-chain alcohol dehydrogenases

- Moreover, 12 primer sets for the short-chain ADHs, which are directed against the three clusters of this group, were finally assayed. The template used was DNA which had been isolated from 5 isolates which, owing to their ADH activity,  
30 had reduced either 4-chloroacetophenone or 2-heptanone in the activity screening. The identity of the amino acid sequence tags to known short-chain ADH sequences is between and the vast majority of these sequences exhibiting an identity of less than to published sequences.



Table 1: sequences which were used for the screening of the DNA sequences encoding theac sof the invention

	<u>Name</u>	<u>Sequence 5' → 3'</u>	<u>Direction</u>	<u>Block</u>
5	ADHM1:	AAAGCATGCGGCGTTTGYCAYACNGA	Forward	A
	ADHM2:	CCAATGTTTCATCGCTTGATATGBNGTRATNCC	Reverse	C
	ADHM3:	TGCGGCGTCTGCCAYACBGA	Forward	A
	ADHM4:	GCTTCAGGGCGTGCTAGGEVGTVAIRCC	Reverse	C
10	ADHM5:	GCGGCGTCTGCCACWCSGA	Forward	A
	ADHM6:	GCTTCAGGGCCTGGTAGGBSGTSAYSCC	Reverse	C
	ADHM7:	AGCCTGCGGCGTCTGYCAYWCBGA	Forward	A
	ADHM8:	GCTTCAGCGCCTGGTAGGBSGTSAYNCC	Reverse	C
	ADHM9:	GCAGCTTGCGGCATGTGYCAYACNGA	Forward	A
15	ADHM10:	GCCCAAGCCGGTCGTAAYNCCRCANCC	Reverse	C
	ADHM11:	GGCCTGCGGCATGTGYCAYACBGA	Forward	A
	ADHM12:	CCCAAGCCGGTCGTGAYRMMRCVCC	Reverse	C
	ADHM13:	CCGGCATGTGCCACACSGA	Forward	A
	ADHM14:	TGGCGGCCAGGCCSAYSSCSCC	Reverse	C
20	ADHM15:	GGCCTCCGGCATGTGYCAYACSGA	Forward	A
	ADHM16:	TGGCGGCCAGGCCSAYNSCNCC	Reverse	C
	ADHM17:	TTAAATGGTGCGGCATTTGYGNCWNGA	Forward	A
	ADHM18:	CAACTTAACAGCCAACATGCCDATNGKNCC	Reverse	D

	ADHM19:	CAAGGTCAAGTGGTGCGGBATYTYGYGG	Forward	A
	ADHM20:	TGACGGCCAACATGCCRATNGKVCC	Reverse	D
	ADHM21:	TGCGGCATCTGCGGSWCSCA	Forward	A
	ADHM22:	CGAACTTGACGGCGAAGAKSCCGATSGKSC	Reverse	D
5	ADHM23:	CAAGGTCAAGTGGTGCGGNATCTGYGG	Forward	A
	ADHM24:	CGGCGAAGATGCCGATSGKNCC	Reverse	D
	ADHM25:	GATTGTTAGAGTTACAGCTACAGCTATTTGYGGNWSNGA	Forward	A
	ADHM26:	TGAACGGCAAACAGGCCNAYNGGNCC	Reverse	D
	ADHM27:	CGCCACCGCCATCTGYGGBWSBGA	Forward	A
10	ADHM28:	GACGGCGAACAGGCCNAYNGGVCC	Reverse	D
	ADHM29:	CACCGCCATCTGCGGSWSSGA	Forward	A
	ADHM30:	GGAGTGGACGGCGAACAKSCCSAYSGGSC	Reverse	D
	ADHM31:	CGCCACCGCCATCTGYGGNWSBGA	Forward	A
	ADHM32:	GACGGCGAACAGGCCSAYSGGNCC	Reverse	D
15	ADHM39	AGAAGAACTGGGCATTATGCCNCCNGGNYT	Forward	A
	ADHM40	TGTATCAATTGTCGGTTGATAGCCNACRAARTCNA	Reverse	D
	ADHM41	ACAACGTGGTCGTGTACGGNCCNTGGGG	Forward	
	ADHM42	GATGGTGGGCTGGTAGCCNACRAARTCNA	Reverse	
	ADHM43	GACAACGTCGTCTACGGNCCNTGGGG	Forward	
20	ADHM44	AGCGCTTGATGGCGTGR TANGNGT	Reverse	
	ADHM45	GACAACGTCGTCTACGGNCCNTGGGG	Forward	
	ADHM46	GATGGTCGGCTGGTAGCCNACRAARTCNA	Reverse	

ADHS1: \_\_\_\_\_

**What is claimed is:**

1. A peptide which has the biological activity of an dependent  
 acid and which comprises or has one of the followinga  
 5 acid csth the sce of the sce of the sce of r a sce which is  
 at least cal to the sce of , the sce of the sce of r a  
 sce which is at least cal to the sce of , the sce of a  
 sce which is at least identical to the sce of , the sce  
 of a sequence which is at least cal to the sce of, the  
 10 sce of a sce which is at least cal to the sce of , the  
 sce of a sce which is at least cal to the sce of , the  
 sce of a sce which is at least cal to the sce of S, the  
 sce of r a sce which is at least cal to the sce of ,  
 the sce of a sce which is at least cal to the sce of  
 15 S, the sce of a sce which is at least cal to the sce of  
 , the sce of a sce which is at least cal to the sce of,  
 the sce of a sce which is at least cal to the sce of,  
 the sce of a sce which is at least cal to the sce of,  
 the sce of a sce which is at least cal to the sce of,  
 20 the sce of a sce which is at least cal to the sce of ,  
 the sce of a sce which is at least cal to the sce of,  
 the sce of a sequence which is at least cal to the sce  
 of, the sce of a sce which is at least cal to the sce  
 of, the sce of a sce which is at least cal to the sce  
 25 of the sce of a sce which is at least cal to the sce  
 of the sce of a sce which is at least cal to the sce of,  
 the sce of a sce which is at least cal to the sequence  
 of the sce of a sce which is at least cal to the sce  
 of the sce of a sce which is at least cal to the sce of,  
 30 the sce of a sce which is at least cal to the sce of  
 the sce of a sce which is at least 70 cal to the sce  
 of the sce of r a sce which is at least cal to the sce  
 of, the sce of a sce which is at least cal to the sce  
 of the sce of r a sce which is at least cal to the sce  
 35 of the sce of the sce of  
 2. A nucleic acid molecule, which encodes the polypeptide  
 as claimed in claim 1.

- 37 -

3. A nucleic acid molecule, which is complementary to the nucleic acid molecule as claimed in claim 2.
4. A vector, which comprises the nucleic acid molecule as claimed in claim 2 or 3.
5. A nonhuman host, which comprises the polypeptide as claimed in claim 1 or the nucleic acid molecule as claimed in claim 2 or 3 or the vector as claimed in claim 4.
6. The host as claimed in claim 5, which is a cell.
7. The host as claimed in claim 5, which is a transgenic nonhuman animal.
8. The host as claimed in any of claims 5 to 7, which has a further dehydrogenase suitable for cofactor regeneration or a nucleic acid molecule encoding said dehydrogenase.
9. The host as claimed in claim 8, in which the dehydrogenase suitable for cofactor regeneration is a formate dehydrogenase or a glucose dehydrogenase.
10. A reaction system, comprising an organic compound which is a substrate of a dehydrogenase, the polypeptide as claimed in claim 1, the vector as claimed in claim 4 or the host as claimed in any of claims 5 to 9 and, where appropriate, a cofactor for the polypeptide as claimed in claim 1.
11. The reaction system as claimed in claim 10, in which the organic compound which is a substrate of a dehydrogenase is a carbonyl compound.
12. The reaction system as claimed in claim 11, in which the carbonyl compound is an aldehyde or a ketone.

- 38 -

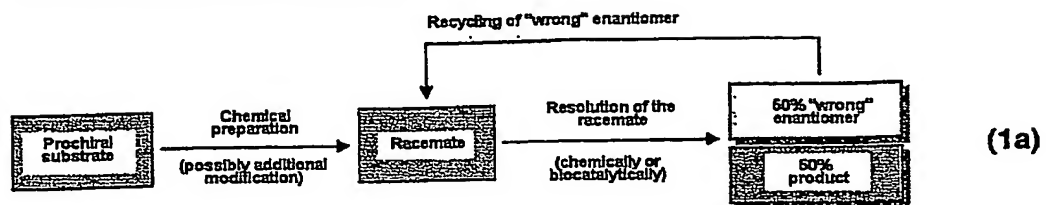
13. The reaction system as claimed in claim 12, in which the ketone is an asymmetrically substituted ketone.
14. The reaction system as claimed in claim 10, in which the organic compound which is a substrate of a dehydrogenase is an alcohol.
15. The reaction system as claimed in claim 14, in which the alcohol is a primary alcohol or a chiral secondary alcohol.
16. The reaction system as claimed in any of claims 10 to 15, in which the cofactor is .
17. A process for preparing the polypeptide as claimed in claim 1 or of a polypeptide encoded by the nucleic acid molecule as claimed in claim 2, which process comprises growing the host as claimed in any of claims 5 to 9 and isolating said polypeptide.
18. A process for preparing the polypeptide as claimed in claim 1 or of a polypeptide encoded by the nucleic acid molecule as claimed in claim 2, which process comprises isolating said polypeptide from a body fluid or tissue sample of the host as claimed in any of claims 7 to 9.
19. A process for an organic compound which is a product of a dehydrogenase, which process comprises reacting an organic compound which is a substrate of a dehydrogenase with the polypeptide as claimed in claim 1, the host as claimed in any of claims 5 to 9 or by means of the reaction system as claimed in any of claims 10 to 16.

20. The process as claimed in claim 19, which further comprises the step of isolating the product of the reaction.
- 5 21. The process as claimed in claim 20, which further comprises processing the product to give a medicament.
22. The process as claimed in claim 20, further comprising the step of processing the product to give a secondary  
10 product.
23. The process as claimed in claim 22, further comprising the step of formulating the secondary product in the preparation of a medicament.  
15
24. The process as claimed in any of claims 19 to 23, in which the product is an enantiomerically pure alcohol.
25. A ligand, which specifically binds the polypeptide as  
20 claimed in claim 1, which ligand is neither a substrate or cofactor of said polypeptide nor a product generated therefrom.
26. The ligand as claimed in claim 25, which is an  
25 antibody or a fragment or derivative thereof, an aptamer, or a low-molecular weight substance.
27. A primer, having a sequence depicted in Table 1.
- 30 28. A primer pair, having sequences depicted in Table 1, with the first primer of said primer pair serving as a forward primer and the second primer of said primer pair serving as a reverse primer to amplify a DNA sequence.
- 35 29. A kit, comprising
- (a) the polypeptide as claimed in claim 1;

- 5 (b) the nucleic acid molecule as claimed in claim 2 or 3;
- (c) the vector as claimed in claim 4;
- (d) the host as claimed in any of claims 6 to 9;
- 10 (e) the reaction system as claimed in any of claims 10 to 16;
- (f) the ligand as claimed in claim 25 or 26;
- (g) at least one primer as claimed in claim 27; and/or
- 15 (h) at least one primer pair as claimed in claim 28.

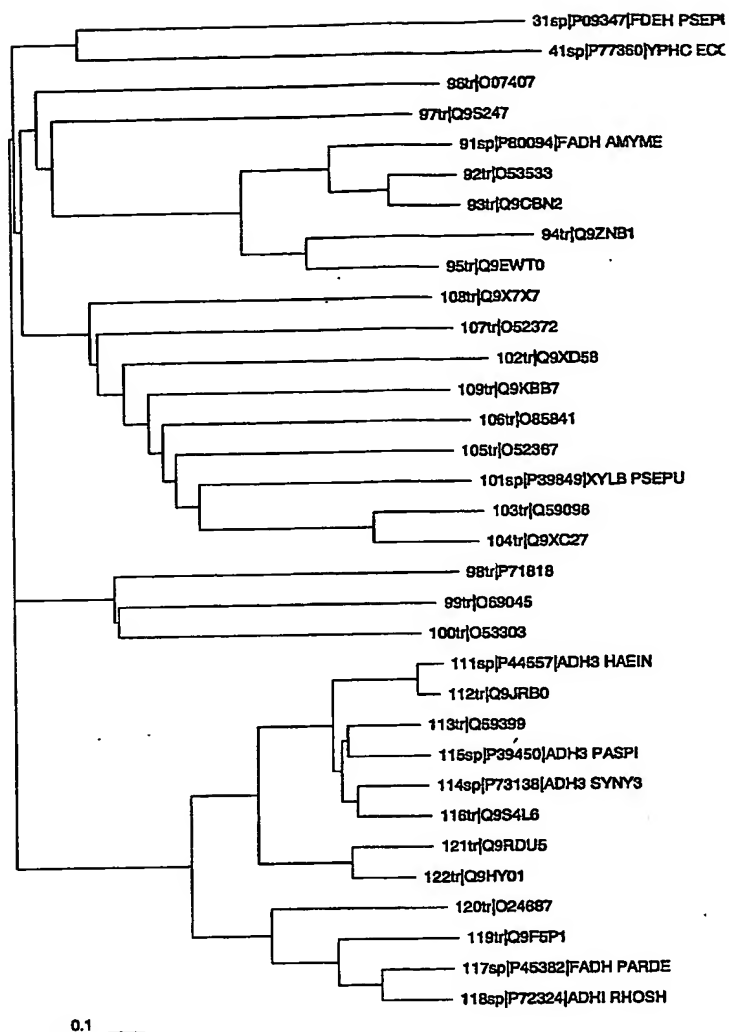


- Prior art via resolution of the racemate: at least 3-4 steps



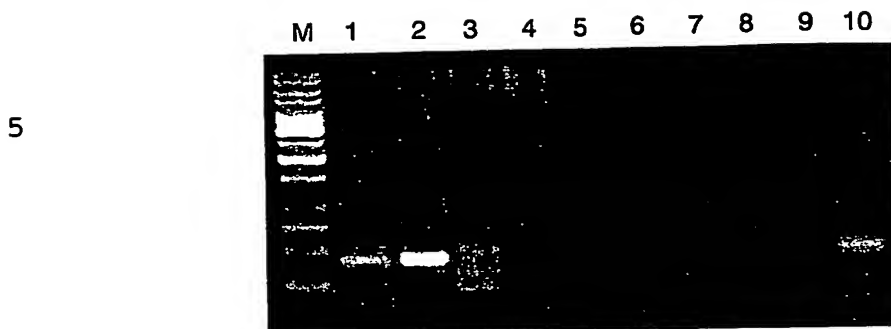
- Biocatalytic and sustainable concept: Asymmetrical biocatalysis





**Fig. 2.**

Overview of cluster 2 (= primer group 2), based on 33 sequences



10 **Fig. 3**

PCR typing with primer group 2, using various pools

Lane occupation: M: Marker 1 kb DNA ladder; Lane 1: Pool 1 with primers ADHM9 + 10; Lane 2: Pool 1 with primers ADHM11 + 12; Lane 3: Pool 1 with primers ADHM13 + 14; Lane 4: Pool 1 with primers ADHM15 + 16; Lane 5: Pool 2 with primers ADHM 9 + 10; Lane 6: Pool 2 with primers ADHM11 + 12; Lane 7: Pool 2 with primers ADHM13 + 14; Lane 8: Pool 2 with primers ADHM15 + 16; Lane 9: Pool 3 with primers ADHM9 + 10; Lane 10: Pool 3 with primers ADHM11 + 12

15

## Sequence Listing

5     <110>   Degussa AG  
       <120>   Novel alcohol dehydrogenases  
       <130>   S-IPM-PAT/Dr. Re-kö - K1419 EP  
 10    <160>   68  
       <170>   PatentIn version 3.1  
       <210>   1  
 15    <211>   162  
       <212>   PRT  
 20    <213>   unknown  
       <220>  
       <221>   source  
 25    <223>   ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
               ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
               ZF0050330= Bacillus; ZF0051303= Bacterium; ZF0051337=  
               Methylobacter; ZF0051321= Bacterium; ZF0050782= Lactobacillus  
 30    Rhodococcus; ZF0050544= Phyllobacterium rubiacearum; ZF0002852=  
               Rhodococcus; ZF0050310= Arthrobacter paraffineus; ZF0002862=  
               Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0002031=  
               Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
               Streptomyces; ZF0002437= Streptomyces; ZF0003712=  
               Micromonospora; ZF0003765= Streptomyces; ZF0051305= Bacterium;  
 35    ZF0003513= Actinomyces; ZF0050993= Kocuria; ZF0002018=  
               Streptomyces; ZF0003767= Actinomyces; ZF0002332= Streptomyces  
               diastatochromogenes; ZF0003768= Actinomyces; ZF0002379=  
               Streptomyces coelestis; ZF0002351= Nonomuraea roseoviolacea;  
               ZF0003769= Actinomyces;  
 40    <400>   1  
           Gly Pro Trp Gly Cys Gly Asn Cys Trp His Cys Ser Gln Gly Leu Glu  
           1                   5                   10                   15  
 45    Asn Tyr Cys Ser Arg Ala Gln Glu Leu Gly Ile Asn Pro Pro Gly Leu  
           20                   25                   30  
 50    Gly Ala Pro Gly Ala Leu Ala Glu Phe Met Ile Val Asp Ser Pro Arg  
           35                   40                   45  
 55    His Leu Val Pro Ile Gly Asp Leu Asp Pro Val Lys Thr Val Pro Leu  
           50                   55                   60

5 Thr Asp Ala Gly Leu Thr Pro Tyr His Ala Ile Lys Arg Ser Leu Pro  
 65 70 75 80  
 Lys Leu Arg Gly Gly Ser Tyr Ala Val Val Ile Gly Thr Gly Gly Leu  
 85 90 95  
 10 Gly His Val Ala Ile Gln Leu Leu Arg His Leu Ser Ala Ser Thr Val  
 100 105 110  
 15 Ile Ala Leu Asp Val Ser Ala Asp Lys Leu Glu Leu Ala Thr Lys Val  
 115 120 125  
 20 Gly Ala His Glu Val Val Leu Ser Asp Lys Asp Ala Ala Glu Asn Val  
 130 135 140  
 25 Arg Lys Ile Thr Gly Ser Gln Gly Ala Ala Leu Val Leu Asp Phe Val  
 145 150 155 160  
 Gly Tyr  
 30 <210> 2  
 <211> 128  
 35 <212> PRT  
 <213> unknown  
 <220>  
 40 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
 45 ZF0050330= Bacillus; ZF0051303= Bacterium; ZF0051337=  
 Methylobacter; ZF0051321= Bacterium; ZF0050782= Lactobacillus  
 bulgaricus; ZF0050544= Phyllobacterium rubiacearum; ZF0002852=  
 Rhodococcus; ZF0050310= Arthrobacter paraffineus; ZF0002862=  
 Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0002031=  
 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 50 Streptomyces; ZF0002437= Streptomyces; ZF0003712=  
 Micromonospora; ZF0003765= Streptomyces; ZF0051305= Bacterium;  
 ZF0003513= Actinomyces; ZF0050993= Kocuria; ZF0002018=  
 Streptomyces; ZF0003767= Actinomyces; ZF0002332= Streptomyces  
 diastatochromogenes; ZF0003768= Actinomyces; ZF0002379=  
 55 Streptomyces coelestis; ZF0002351= Nonomuraea roseoviolacea;  
 ZF0003769= Actinomyces;

&lt;400&gt; 2

Gly Pro Trp Gly Cys Gly Asn Cys Trp His Cys Ser Gln Gly Leu Glu  
 1 5 10 15  
 5 Asn Tyr Cys Ser Arg Ala Gln Glu Leu Gly Ile Asn Pro Pro Gly Leu  
 20 25 30  
 10 Gly Ala Pro Gly Ala Leu Ala Glu Phe Met Ile Val Asp Ser Pro Arg  
 35 40 45  
 15 His Leu Val Pro Ile Gly Asp Leu Asp Pro Val Lys Thr Val Pro Leu  
 50 55 60  
 20 Thr Asp Ala Gly Leu Thr Pro Tyr His Ala Ile Lys Arg Ser Leu Pro  
 65 70 75 80  
 Lys Leu Arg Gly Gly Ser Tyr Ala Val Val Ile Gly Thr Gly Gly Leu  
 85 90 95  
 25 Gly His Val Thr Ile Gln Leu Leu Arg His Leu Ser Ala Ala Thr Val  
 100 105 110  
 30 Ile Ala Leu Asp Val Ser Ala Asp Lys Leu Glu Leu Ala Thr Lys Val  
 115 120 125

&lt;210&gt; 3

35

&lt;211&gt; 162

&lt;212&gt; PRT

40

&lt;213&gt; unknown

&lt;220&gt;

&lt;221&gt; source

45

<223> ZF0050286= *Corynebacterium hoagii*

&lt;400&gt; 3

Gly Pro Trp Gly Cys Gly Arg Cys Trp His Cys Ala Gln Gly Leu Glu  
 1 5 10 15  
 50 Asn Tyr Cys Ser Arg Ala Arg Glu Leu Gly Ile Ala Pro Pro Gly Leu  
 20 25 30  
 55 Gly Ala Pro Gly Ala Ile Ala Glu Tyr Met Ile Val Asp Ser Pro Arg  
 35 40 45

His Leu Val Pro Ile Gly Asp Leu Asp Pro Val Thr Thr Val Pro Leu  
 50 55 60

5 Thr Asp Ala Gly Leu Thr Pro Tyr His Ala Ile Lys Arg Ser Leu Gly  
 65 70 75 80

10 Lys Leu Arg Ala Gly Ser Tyr Ala Val Val Ile Gly Thr Gly Gly Leu  
 85 90 95

15 Gly His Val Gly Ile Gln Leu Leu Arg His Leu Ser Pro Ala Arg Ile  
 100 105 110

20 Ile Ala Leu Asp Val Asn Asp Glu Lys Leu Ala Phe Ala Arg Glu Val  
 115 120 125

Gly Ala His Glu Thr Val Leu Ser Asn Ala Asp Ala Ala Ala Asn Val  
 130 135 140

25 Arg Lys Ile Thr Gly Ser Ala Gly Ala Ala Leu Val Leu Asp Phe Val  
 145 150 155 160

30 Gly Tyr

35 <210> 4  
 <211> 161  
 <212> PRT  
 40 <213> unknown  
 <220>

45 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 4

50 Gly Pro Trp Gly Cys Gly Ser Cys Trp His Cys Ser Gln Gly Leu Glu  
 1 5 10 15

55 Asn Tyr Cys Ser Arg Ala Lys Glu Leu Gly Ile Asn Pro Pro Gly Leu  
 20 25 30

Gly Ala Pro Gly Ala Leu Ala Glu Phe Met Ile Val Asp Ser Pro Arg

	35	40	45
5	His Leu Val Pro Ile Gly Asp Leu Asp Pro Val Lys Thr Val Pro Leu 50 55 60		
	Thr Asp Ala Gly Leu Thr Pro Tyr His Ala Ile Lys Arg Ser Leu Pro 65 70 75 80		
10	Lys Leu Arg Gly Gly Ala Tyr Ala Val Val Ile Gly Thr Gly Gly Leu 85 90 95		
15	Gly His Val Ala Ile Gln Leu Leu Arg His Leu Ser Ala Ala Thr Val 100 105 110		
20	Ile Ala Leu Asp Val Ser Ala Asp Lys Leu Val Leu Ala Thr Lys Val 115 120 125		
	Gly Ala His Glu Val Val Leu Ser Asp Lys Asp Ala Ala Glu Asn Val 130 135 140		
25	Arg Arg Ile Thr Gly Ser Gln Gly Ala Ala Leu Val Leu Asp Phe Val 145 150 155 160		
30	Gly		
35	<210> 5		
	<211> 70		
	<212> PRT		
40	<213> unknown		
	<220>		
	<221> source		
45	<223> ZF0004210= Actinomyces; ZF0004212= Actinomyces; ZF0004211=		
	Actinomyces; ZF0003860= Actinomyces; ZF0004218= Actinomyces;		
	ZF0003868= Actinomadura; ZF0004213= Actinomyces; ZF0003876=		
	Actinomyces; ZF0003866= Actinomyces; ZF0003864= Actinomyces;		
50	ZF0003862= Actinomadura; ZF0003869= Actinomyces; ZF0003867=		
	Actinomadura; ZF0004216= Actinomyces; ZF0004235= Actinomyces;		
	ZF0004209= Actinomadura; ZF0004214= Actinomyces; ZF0003871=		
	Actinomyces; ZF0004063= Actinomadura; ZF0004052= Actinomadura;		
	ZF0006405= Streptomyces; ZF0003865= Actinomadura; ZF0004047=		
	Actinomadura; ZF0004070= Actinomyces; ZF0004085= Actinomyces;		
55	ZF0004217= Actinomyces; ZF0004089= Actinomadura; ZF0004090=		
	Actinomadura; ZF0006138= Streptomyces; ZF0004236= Actinomadura;		
	ZF0051203= Bacterium;		
	<400> 5		



Gly Pro Trp Gly Cys Gly Thr Cys Val Lys Cys Ala Glu Gly Lys Glu  
 1 5 10 15  
 5  
 Asn Tyr Cys Leu Arg Ala Lys Glu Leu Gly Ile Ala Pro Pro Gly Leu  
 20 25 30  
 10 Gly Ser Pro Gly Ala Met Ala Glu Tyr Met Ile Val Asp Asp Pro Arg  
 35 40 45  
 15 His Leu Val Pro Leu Gly Gly Leu Asp Pro Val Gln Ala Val Pro Leu  
 50 55 60  
 Thr Asp Ala Gly Leu Thr  
 65 70  
 20  
 <210> 6  
 <211> 94  
 25  
 <212> PRT  
 <213> unknown  
 30  
 <220>  
 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
 35 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
 40 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
 45 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 Actinomyces; ZF0003535= Actinomyces;  
 <400> 6  
 50  
 Cys His Thr Asp His His Ile Val Thr Gly Ala Thr Pro Met Pro Ser  
 1 5 10 15  
 55 Phe Pro Val Met Gly Gly His Glu Gly Ser Gly Val Ile Thr Lys Leu  
 20 25 30  
 Gly Pro Glu Val Lys Gly Leu Glu Val Gly Asp His Val Val Leu Ser

35 40 45  
 5 Phe Ile Pro Ala Cys Gly Thr Cys Pro Ala Cys Ser Ala Gly His Gln  
 50 55 60  
 10 Asn Leu Cys Asp Leu Gly Met Gly Leu Leu Ser Gly Gln Ala Ile Ser  
 65 70 75 80  
 Asp Gly Thr Tyr Arg Ile Gln Ala Arg Gly Glu Asn Val Ile  
 85 90  
 15 <210> 7  
 <211> 92  
 20 <212> PRT  
 <213> unknown  
 <220>  
 25 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
 30 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 35 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
 40 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 Actinomyces; ZF0003535= Actinomyces;  
 <400> 7  
 45 Cys His Thr Asp Asp His Ala Val Thr Gly Asp Leu Ala Val Pro Leu  
 1 5 10 15  
 50 Pro Val Ile Gly Gly His Glu Gly Ala Gly Ile Val Glu Lys Val Gly  
 20 25 30  
 55 Pro Gly Val Arg Asp Val Glu Val Gly Asp His Val Val Leu Ser Phe  
 35 40 45  
 Ile Pro Ser Cys Gly Arg Cys Arg Trp Cys Ala Val Gly Gln Ser Asn  
 50 55 60

5 Leu Cys Asp Leu Gly Ala Ile Leu Met Ala Gly Ala Gln Val Asp Gly  
 65 70 75 80

10 Thr Tyr Arg Ala Thr Ala Arg Gly His Asp Val Gly  
 85 90

15 <210> 8  
 <211> 92  
 <212> PRT  
 <213> unknown  
 <220>

20 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=

25 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=

30 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=

35 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 Actinomyces; ZF0003535= Actinomyces;

40 <400> 8

Cys His Thr Asp Asp His Ala Val Thr Gly Asp Leu Ala Val Pro Leu  
 1 5 10 15

45 Pro Val Ile Gly Gly His Glu Gly Ala Gly Ile Val Glu Lys Val Gly  
 20 25 30

50 Pro Gly Val Arg Asp Val Glu Val Gly Asp His Val Val Leu Ser Phe  
 35 40 45

55 Ile Pro Ser Cys Gly Arg Cys Arg Trp Cys Ala Val Gly Gln Ser Asn  
 50 55 60

Leu Cys Asp Leu Gly Ala Ile Leu Met Ala Gly Ala Gln Val Asp Gly  
 65 70 75 80

Thr Tyr Arg Ala Thr Ala Arg Gly His Asp Val Gly  
 85 90

5 <210> 9  
 <211> 92

10 <212> PRT  
 <213> unknown  
 <220>

15 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=

20 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434= Streptomyces;  
 ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331= Actinoplanes philippinensis;  
 ZF0002441= Streptomyces;

25 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240= Bacterium;  
 ZF0002333= Rhodococcus erythropolis; ZF0003713= Micromonospora; ZF0004980= Streptomyces;  
 ZF0004821= Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396= Actinoplanes;  
 ZF0003781= Actinomyces; ZF0003512= Actinomyces; ZF0006093= Streptomyces;  
 ZF0006103= Streptomyces; ZF0006087= Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 30 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538= Actinomyces;  
 ZF0003535= Actinomyces;

<400> 9

35 Cys His Thr Asp Asp His Ala Val Thr Gly Asp Leu Ala Val Pro Leu  
 1 5 10 15

40 Pro Val Ile Gly Gly His Glu Gly Ala Gly Ile Val Glu Lys Val Gly  
 20 25 30

45 Pro Gly Val Arg Asp Val Glu Val Gly Asp His Val Val Leu Ser Phe  
 35 40 45

50 Ile Pro Ser Cys Gly Arg Cys Arg Trp Cys Ala Val Gly Gln Ser Asn  
 50 55 60

Leu Cys Asp Leu Gly Ala Ile Leu Met Ala Gly Ala Gln Val Asp Gly  
 65 70 75 80

55 Thr Tyr Arg Ala Thr Ala Arg Gly His Asp Val Gly  
 85 90

<210> 10  
 <211> 92  
 5 <212> PRT  
 <213> unknown  
 <220>  
 10 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
 15 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 20 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
 25 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 Actinomyces; ZF0003535= Actinomyces;  
 <400> 10  
 30 Cys His Thr Asp Asp His Ala Val Thr Gly Asp Leu Ala Val Pro Leu  
 1 5 10 15  
 35 Pro Val Ile Gly Gly His Glu Gly Ala Gly Ile Val Glu Lys Val Gly  
 20 25 30  
 Pro Gly Val Arg Asp Val Glu Val Gly Asp His Val Val Leu Ser Phe  
 40 35 40 45  
 Ile Pro Ser Cys Gly Arg Cys Arg Trp Cys Ala Val Gly Gln Ser Asn  
 45 50 55 60  
 Leu Cys Asp Leu Gly Ala Ile Leu Met Ala Gly Ala Arg Val Asp Gly  
 65 70 75 80  
 50 Thr Tyr Arg Ala Thr Ala Arg Gly His Asp Val Gly  
 85 90  
 55 <210> 11  
 <211> 92  
 <212> PRT

<213> unknown  
 <220>  
 5 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
 10 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
 15 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
 20 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 Actinomyces; ZF0003535= Actinomyces;  
 <400> .11  
 25 Cys His Thr Asp Asp His Ala Val Thr Gly Asp Leu Ala Val Pro Leu  
 1 5 10 15  
 30 Pro Val Ile Gly Gly His Glu Gly Ala Gly Ile Val Glu Lys Val Gly  
 20 25 30  
 35 Pro Gly Val Arg Asp Val Glu Val Gly Asp His Val Val Leu Ser Phe  
 35 40 45  
 40 Ile Pro Ser Cys Gly Arg Cys Arg Trp Cys Ala Val Gly Gln Ser Asn  
 50 55 60  
 45 Leu Cys Asp Leu Gly Ala Ile Leu Met Ala Gly Ala Gln Val Asp Gly  
 65 70 75 80  
 50 Thr Tyr Arg Ala Thr Ala Arg Gly His Asp Val Gly  
 85 90  
 55 <213> unknown  
 <220>  
 <221> source

<223> ZF0050310= *Arthrobacter paraffineus*

<400> 12

5  
Cys His Thr Asp Leu Phe Thr Lys Ser Val Leu Pro Glu Arg Leu Gly  
1 5 10 15

10 Pro Cys Val Phe Gly His Glu Gly Ala Gly Val Val Glu Ala Val Gly  
20 25 30

15 Ser Ser Ile Asp Ser Ile Ala Pro Gly Asp His Val Leu Leu Ser Tyr  
35 40 45

20 Arg Ser Cys Gly Val Cys Arg Gln Cys Leu Ser Gly His Arg Ala Tyr  
50 55 60

Cys Glu Ser Ser His Gly Leu Asn Ser Ser Gly Ala Arg Thr Asp Gly  
65 70 75 80

25 Ser Thr Pro Val Arg Arg Ser Gly Thr Pro Ile Arg Ser  
85 90

30 <210> 13

<211> 93

<212> PRT

35

<213> unknown

<220>

40

<221> source

<223> ZF0002333=*Rhodococcus erythropolis*

<400> 13

45 Cys His Thr Asp Leu Phe Thr Lys Thr Val Leu Pro Glu Lys Leu Gly  
1 5 10 15

50 Pro Cys Val Phe Gly His Glu Gly Ala Gly Val Val Gln Ala Val Gly  
20 25 30

55 Ser Ser Ile Asp Asn Ile Ala Ala Gly Asp His Val Leu Leu Ser Tyr  
35 40 45

Arg Ser Cys Gly Val Cys Arg Gln Cys Leu Ser Asp His Arg Ala Tyr  
50 55 60

Cys Glu Ser Ser His Gly Leu Asn Ser Ser Gly Ala Arg Thr Asp Gly  
 65 70 75 80

5 Ser Thr Pro Val Arg Arg Asn Gly Thr Pro Ile Arg Ser  
 85 90

10 <210> 14  
 <211> 120  
 <212> PRT  
 15 <213> unknown  
 <220>  
 20 <221> source  
 <223> ZF0051303= Bacterium; ZF0051337= Methylobacter; ZF0002862=  
 Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0051305=  
 Bacterium; ZF0003513= Actinomyces; ZF0002351= Nonomuraea  
 25 roseoviolacea; ZF0003769= Actinomyces; ZF0002017= Streptomyces;  
 ZF0051306= Bacterium; ZF0002016= Streptomyces; ZF0003504=  
 Actinomyces; ZF0006073= Streptomyces; ZF0003770= Actinomyces;  
 ZF0002352= Actinoplanes italicus; ZF0002378= Streptomyces  
 aureomonopodiales; ZF0006089= Streptomyces; ZF0006106=  
 Streptomyces; ZF0051325= Bacterium; ZF0006108= Streptomyces;  
 30 ZF0002440= Streptomyces; ZF0051302= Bacterium; ZF0003532=  
 Actinomyces; ZF0003548= Nocardiaform;  
 <400> 14

35 Cys Gly Thr Asp Arg Glu Ile Ala Ser Gly Ile Tyr Gly Trp Ala Pro  
 1 5 10 15

40 Pro Gly Arg Glu His Leu Val Leu Gly His Glu Ser Leu Gly Arg Val  
 20 25 30

45 Arg Thr Ala Pro Asp Gly Ser Gly Phe Ala Ala Gly Asp Leu Val Val  
 35 40 45

Gly Ile Val Arg Arg Pro Asp Pro Val Pro Cys Gly Ala Cys Ala His  
 50 55 60

50 Gly Glu Phe Asp Met Cys Arg Asn Gly Glu Tyr Val Glu Arg Gly Ile  
 65 70 75 80

55 Lys Gln Ile Asp Gly Tyr Gly Ser Thr Ser Trp Val Val Asp Ala Asp  
 85 90 95

Tyr Thr Val Lys Leu Asp Pro Ala Leu Thr Glu Val Gly Val Leu Met



	100	105	110
	Glu Pro Thr Thr Val Leu Gly Gln		
5	115	120	
	<210> 15		
	<211> 140		
10	<212> PRT		
	<213> unknown		
	<220>		
15	<221> source		
	<223> ZF0051303= Bacterium; ZF0051337= Methylomonas; ZF0002862= Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0051305= Bacterium; ZF0003513= Actinomyces; ZF0002351= Nonomuraea roseoviolacea; ZF0003769= Actinomyces; ZF0002017= Streptomyces; ZF0051306= Bacterium; ZF0002016= Streptomyces; ZF0003504= Actinomyces; ZF0006073= Streptomyces; ZF0003770= Actinomyces; ZF0002352= Actinoplanes italicus; ZF0002378= Streptomyces aureomonopodiales; ZF0006089= Streptomyces; ZF0006106= Streptomyces; ZF0051325= Bacterium; ZF0006108= Streptomyces; ZF0002440= Streptomyces; ZF0051302= Bacterium; ZF0003532= Actinomyces; ZF0003548= Nocardiaform;		
	<400> 15		
30	Cys Gly Thr Asp Leu His Ile Arg Ser Trp Asp Gly Trp Ala Gln Lys		
	1 5 10 15		
35	Thr Ile Ala Thr Pro Leu Thr Leu Gly His Glu Phe Val Gly Glu Val		
	20 25 30		
40	Val Glu Thr Gly Arg Asp Val Thr Asp Ile Gln Val Gly Asp Leu Val		
	35 40 45		
45	Ser Gly Glu Gly His Leu Val Cys Gly Lys Cys Arg Asn Cys Leu Ala		
	50 55 60		
	Gly Arg Arg His Leu Cys Arg Ala Thr Val Gly Leu Gly Val Gly Arg		
	65 70 75 80		
50	Asp Gly Ala Phe Ala Glu Tyr Val Val Leu Pro Ala Ser Asn Val Trp		
	85 90 95		
55	Val His Arg Val Pro Val Asp Leu Asp Val Ala Ala Ile Phe Asp Pro		
	100 105 110		
	Phe Gly Asn Ala Val His Thr Ala Leu Ser Phe Pro Leu Val Gly Glu		

	115	120	125
5	Asp Val Leu Val Thr Gly Ala Gly Thr Ile Gly Ile 130 135 140		
	<210> 16		
	<211> 138		
10	<212> PRT		
	<213> unknown		
	<220>		
15	<221> source		
	<223> ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;		
	ZF0050330= Bacillus, ZF0002852= Rhodococcus; ZF0050310=		
20	Arthrobacter paraffineus; ZF0002437= Streptomyces; ZF0003712=		
	Micromonospora; ZF0003765= Streptomyces; ZF0002332=		
	Streptomyces diatsatochromogenes; ZF0003768= Actinomyces;		
	ZF0002379= Streptomyces coelestis; ZF0002443= Streptomyces;		
	ZF0002442= Streptomyces; ZF0002436= Streptomyces; ZF0050994=		
25	Bacterium; ZF0050992= Bacterium; ZF0050442= Bacterium;		
	ZF0002049= Streptomyces; ZF0006069= Streptomyces; ZF0006075=		
	Streptomyces; ZF0004724= Nocardiaform; ZF0002392= Actinoplanes		
	nipponensis; ZF0002356= Actinoplanes brasiliensis; ZF0003501=		
	Actinomyces; ZF0051322= Bacterium; ZF0006078= Streptomyces;		
30	ZF0006092= Streptomyces; ZF0006090= Streptomyces; ZF0006084=		
	Streptomyces; ZF0006068= Streptomyces; ZF0050284= Rhodococcus;		
	ZF0050028= Agrobacterium tumefaciens; ZF0003540= Actinomyces;		
	ZF0003528= Actinomyces; ZF0003529= Actinomyces;		
	<400> 16		
35	Gly Leu Thr Ile Gly His Glu Pro Val Gly Val Ile Glu Lys Leu Gly 1 5 10 15		
40	Ser Ala Val Thr Gly Tyr Arg Glu Gly Gln Arg Val Ile Ala Gly Ala 20 25 30		
45	Ile Cys Pro Asn Phe Asn Ser Tyr Ala Ala Gln Asp Gly Ala Pro Ser 35 40 45		
50	Gln Asp Gly Ser Tyr Leu Val Ala Ser Gly Ala Cys Gly Cys His Gly 50 55 60		
55	Tyr Arg Ala Thr Ala Gly Trp Arg Phe Gly Asn Ile Ile Asp Gly Ala 65 70 75 80		
	Gln Ala Glu Tyr Leu Leu Val Pro Asp Ala Gln Gly Asn Leu Ala Pro 85 90 95		

16

Val Pro Asp Asn Leu Ser Asp Glu Gln Val Leu Met Cys Pro Asp Ile  
 100 105 110

5

Met Ser Thr Gly Phe Lys Gly Ala Glu Asn Ala His Ile Arg Ile Gly  
 115 120 125

10 Asp Thr Val Ala Val Phe Ala Gln Gly Pro  
 130 135

<210> 17

15

<211> 144

<212> PRT

20 <213> unknown

<220>

<221> source

25 <223> ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
 ZF0050330= Bacillus, ZF0002852= Rhodococcus; ZF0050310=  
 Arthrobacter paraffineus; ZF0002437= Streptomyces; ZF0003712=  
 Micromonospora; ZF0003765= Streptomyces; ZF0002332=  
 Streptomyces diastatochromogenes; ZF0003768= Actinomyces;  
 30 ZF0002379= Streptomyces coelestis; ZF0002443= Streptomyces;  
 ZF0002442= Streptomyces; ZF0002436= Streptomyces; ZF0050994=  
 Bacterium; ZF0050992= Bacterium; ZF0050442= Bacterium;  
 ZF0002049= Streptomyces; ZF0006069= Streptomyces; ZF0006075=  
 Streptomyces; ZF0004724= Nocardiaform; ZF0002392= Actinoplanes  
 35 nipponensis; ZF0002356= Actinoplanes brasiliensis; ZF0003501=  
 Actinomyces; ZF0051322= Bacterium; ZF0006078= Streptomyces;  
 ZF0006092= Streptomyces; ZF0006090= Streptomyces; ZF0006084=  
 Streptomyces; ZF0006068= Streptomyces; ZF0050284= Rhodococcus;  
 ZF0050028= Agrobacterium tumefaciens; ZF0003540= Actinomyces;  
 40 ZF0003528= Actinomyces; ZF0003529= Actinomyces;

<400> 17

45 Cys Gly Thr Asp Leu His Ile Leu Gly Gly Asp Val Pro Glu Val Thr  
 1 5 10 15

Asp Gly Arg Ile Leu Gly His Glu Ala Val Gly Thr Val Val Glu Val  
 20 25 30

50

Gly Asp Gly Val Gln Thr Leu Ala Pro Gly Asp Arg Val Leu Val Ser  
 35 40 45

55

Cys Val Thr Ala Cys Gly Thr Cys Arg Phe Cys Arg Glu Ser Arg Tyr  
 50 55 60

Gly Gln Cys Leu Gly Gly Gly Gly Trp Ile Leu Gly His Leu Ile Asp  
 65 70 75 80

5 Gly Thr Gln Ala Glu Leu Val Arg Val Pro Tyr Ala Asp Asn Ser Thr  
 85 90 95

10 His Arg Ile Pro Asp Gly Val Ser Asp Glu Gln Met Leu Met Leu Ala  
 100 105 110

15 Asp Ile Leu Pro Thr Ser Tyr Glu Val Gly Val Leu Asn Gly Cys Leu  
 115 120 125

20 Arg Pro Ala Asp Val Val Val Ile Ile Gly Ala Asp Asp Arg Pro Leu  
 130 135 140

<210> 18

25 <211> 73

<212> PRT

<213> unknown

30 <220>

<221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*

35 <400> 18

Val Asp Val Val Val Asp Asn Ala Gly Phe Gly Thr His Gly Ala Phe  
 1 5 10 15

40 Val Asp Glu Asp His Glu Arg Val Thr Ser Glu Ile Gln Leu Asn Ile  
 20 25 30

45 Ala Thr Leu Val Glu Leu Thr His Thr Phe Pro Pro Asp Leu Leu Thr  
 35 40 45

Gly Arg Gly Ala Leu Val Asn Ile Ala Ser Thr Ala Ser Phe Gln Pro  
 50 55 60

5 Thr Pro Gly Met Ala Val Tyr Cys Ala  
 65 70

<210> 19  
 10 <211> 75  
 <212> PRT  
 15 <213> unknown  
 <220>  
 <221> source  
 20 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 19

Val Asp Val Val Val His Asn Ala Gly Phe Gly Thr His Gly Ala Phe  
 25 1 5 10 15

Val Asp Glu Asp Leu Glu Arg Val Thr Ser Glu Ile Gln Leu Asn Ile  
 30 20 25 30

Ala Thr Leu Val Glu Leu Thr His Thr Phe Leu Pro Asp Leu Leu Thr  
 35 35 40 45

Gly Arg Gly Ala Leu Val Asn Ile Ala Ser Thr Ala Ser Phe Gln Pro  
 50 55 60

40 Thr Pro Gly Met Ala Val Tyr Cys Ala Thr Lys  
 65 70 75

<210> 20  
 45 <211> 79  
 <212> PRT  
 50 <213> unknown  
 <220>  
 <221> source  
 55 <223> ZF0003535= *Actinomyces*  
 <400> 20

Arg Val Asp Val Val Val His Asn Ala Ala Ile Thr Gln Lys Ala Thr



<212> PRT  
<213> unknown  
5 <220>  
<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*  
10 <400> 22  
Gln Gly Ile Gly Tyr Ala Thr Ala Lys Arg Leu Ile Ser Leu Gly Ala  
1 5 10 15  
15 Thr Val Ala Ile Gly Asp Ile Asp Glu Ala Thr Leu Ala Arg Ala Ala  
20 25 30  
20 Lys Asp Leu Gly Ile Arg Thr Phe Gly Arg Leu Asp Val Thr Asp Pro  
35 40 45  
25 Ala Ser Phe Phe Asp Phe Leu Asp Thr Val Glu Gly Glu Leu Gly Pro  
50 55 60  
30 Ile Asp Val Leu Ile Asn Asn Ala  
65 70  
<210> 23  
35 <211> 75  
<212> PRT  
<213> unknown  
40 <220>  
<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*  
45 <400> 23  
Gln Arg Ile Gly Leu Glu Ile Ala Arg Thr Phe Ile Lys Glu Gly Ala  
1 5 10 15  
50 Thr Val Val Leu Gly Asp Ile Asn Glu Thr Val Gly Thr Ala Ala Val  
20 25 30  
55 Ala Glu Leu Gly Gly Glu Ser Val Ala Arg Phe Ala Ser Cys Asp Val  
35 40 45

Arg Asp Ser Gly Gln Val Glu Ala Met Leu Asp Leu Ala Glu Ser Ala  
 50 55 60

5 Phe Gly Pro Val Asp Val Met Met Asn Asn Ala  
 65 70 75

10 <210> 24  
 <211> 72  
 <212> PRT

15 <213> unknown  
 <220>  
 <221> source  
 20 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 24

25 Gln Gly Ile Gly Tyr Gln Thr Ala Lys Glu Leu Ile Arg Arg Gly His  
 1 5 10 15

30 Arg Val Ala Ile Gly Asp Ile Asp Glu Ala Arg Ala Lys Glu Thr Ala  
 20 25 30

Ala Glu Leu Gly Val Lys Val Val Thr Arg Leu Asp Val Thr Asp Pro  
 35 35 40 45

35 Asp Ser Phe Lys Asp Phe Leu Asp Leu Val Glu Gly Asp Leu Gly Pro  
 50 55 60

40 Leu Asp Val Leu Ile Asn Asn Ala  
 65 70

45 <210> 25  
 <211> 74  
 <212> PRT

50 <213> unknown  
 <220>  
 <221> source  
 55 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 25



Gly Ile Gly Leu Glu Ile Ala Arg Thr Phe Ile Lys Glu Gly Ala Thr  
 1 5 10 15  
 5 Val Val Leu Gly Asp Ile Asn Glu Thr Val Gly Thr Ala Ala Val Ala  
 20 25 30  
 10 Glu Leu Gly Gly Glu Ser Val Ala Arg Phe Ala Ser Cys Asp Val Arg  
 35 40 45  
 15 Asp Ser Gly Gln Val Glu Ala Met Leu Asp Leu Ala Glu Ser Ala Phe  
 50 55 60  
 20 Gly Pro Val Asp Val Ile Val Asn Asn Ala  
 65 70  
 <210> 26  
 <211> 74  
 <212> PRT  
 <213> unknown  
 30 <220>  
 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 35 <400> 26  
 Ile Gly Leu Glu Ile Ala Arg Thr Phe Ile Lys Glu Gly Ala Thr Val  
 1 5 10 15  
 40 Val Leu Gly Asp Ile Asn Glu Thr Val Gly Thr Ala Ala Val Gly Glu  
 20 25 30  
 45 Leu Gly Gly Glu Ser Val Ala Arg Phe Ala Ser Cys Asp Val Arg Asp  
 35 40 45  
 50 Ser Gly Gln Val Glu Ala Met Leu Asp Leu Ala Glu Ser Ala Phe Gly  
 50 55 60  
 55 Pro Val Asp Val Met Val Asn Asn Ala Gly  
 65 70  
 <210> 27  
 <211> 62

5       <212>   PRT  
       <213>   unknown  
       <220>  
       <221>   source  
 10       <223>   ZF0002333= Rhodococcus erythropolis  
       <400>   27  
  
 15       Val Pro Val Ala Val Val Asp Leu His Ile Glu Ser Ala Lys Glu Thr  
           1                   5                   10                   15  
  
 20       Val Ala Leu Ile Glu Ser Gln Tyr Gly Thr Pro Ala Leu Ala Leu Glu  
                   20                   25                   30  
  
       Ala Asp Val Arg Asp Arg Ala Ala Val Ser Ala Ala Phe Glu Ala Thr  
           35                   40                   45  
  
 25       Val Ala Glu Trp Gly Arg Phe Asp Tyr Leu Val Asn Asn Ala  
           50                   55                   60  
  
 30       <210>   28  
       <211>   74  
       <212>   PRT  
 35       <213>   unknown  
       <220>  
       <221>   source  
       <223>   ZF0002333= Rhodococcus erythropolis  
       <400>   28  
  
 45       Leu Gly Arg Glu Ile Ala Leu Lys Leu Ala Ser Glu Gly Ala Ser Val  
           1                   5                   10                   15  
  
 50       Val Val Asn Asp Leu Asp Pro Glu Pro Ala Ala Gln Thr Glu Arg Asp  
                   20                   25                   30  
  
       Ile Lys Ala Thr Gly Gly Gln Ala Val Ser Cys Val Gly Ser Val Ala  
           35                   40                   45  
 55  
  
       Glu Asp Gly Phe Ala Glu Arg Phe Val Asn Thr Ala Val Glu Ser Phe  
           50                   55                   60

Gly Gly Leu Asp Val Met Val Asn Asn Ala  
 65 70

5  
 <210> 29  
 <211> 76

10  
 <212> PRT  
 <213> unknown  
 <220>

15  
 <221> source  
 <223> ZF0002333= *Rhodococcus erythropolis*  
 <400> 29

20  
 Ala Gly Leu Gly Val Glu Phe Ala His Arg Phe Ala Ala Arg Gly Ala  
 1 5 10 15

25  
 Asn Leu Val Leu Val Ala Arg Arg Ala Asp Arg Leu Glu Ala Leu Ala  
 20 25 30

30  
 Thr Glu Leu Arg Val Ala His Gly Ile Thr Val Thr Val Leu Pro Ala  
 35 40 45

35  
 Asp Leu Ala Ala Pro Gly Val Gly Ala Thr Leu His Gln Glu Leu Thr  
 50 55 60

40  
 Ser Arg Gly Ile Thr Val Thr Ser Leu Ile Asn Asn  
 65 70 75

45  
 <210> 30  
 <211> 72  
 <212> PRT  
 <213> unknown  
 <220>

50  
 <221> source  
 <223> ZF0003535= *Actinomyces*  
 <400> 30

55  
 Pro Ala Asp Gly Tyr Gln Thr Ala Lys Glu Leu Ile Arg Arg Gly His  
 1 5 10 15

25

Arg Val Ala Ile Val Asp Ile Asp Glu Ala Arg Ala Lys Gly Ala Ala  
                     20                    25                    30

5   Ala Glu Leu Gly Val Lys Val Val Thr Arg Leu Asp Val Thr Glu Pro  
                     35                    40                    45

10   Asp Ser Phe Thr Thr Phe Leu Asp Leu Val Glu Arg Glu Leu Gly Pro  
                     50                    55                    60

15   Leu Asp Ile Leu Val Asn Asn Ala  
                     65                    70

<210>   31

<211>   67

20   <212>   PRT

<213>   unknown

25   <220>

<221>   source

<223>   ZF0050310= *Arthrobacter paraffineus*

30   <400>   31

Ala Thr Asp Gly Ala Arg Val Ala Val Val Asp Leu His Ile Glu Ser  
 1                    5                    10                    15

35   Ala Glu Glu Thr Val Ala Leu Ile Glu Ser Gln Tyr Gly Thr Pro Ala  
                     20                    25                    30

40   Leu Ala Leu Glu Ala Asp Val Arg Asp Arg Ala Ala Val Ser Ala Ala  
                     35                    40                    45

45   Phe Glu Ala Thr Val Ala Glu Trp Gly Arg Phe Asp Tyr Leu Val Asn  
                     50                    55                    60

50   Asn Ala Gly  
                     65

<210>   32

<211>   67

55   <212>   PRT

<213>   unknown

<220>  
 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 5 <400> 32  
 Ala Ala Asp Gly Ala Arg Val Ala Val Val Asp Leu His Ile Glu Ser  
 1 5 10 15  
 Ala Lys Glu Thr Val Ala Leu Ile Glu Ser Gln Tyr Gly Thr Pro Ala  
 20 25 30  
 15 Leu Ala Leu Glu Ala Asp Val Arg Asp Arg Ala Ala Val Ser Ala Ala  
 35 40 45  
 Phe Glu Ala Thr Val Ala Glu Trp Gly Arg Phe Asp Tyr Leu Val Asn  
 50 55 60  
 20 Asn Ala Gly  
 65  
 25 <210> 33  
 <211> 348  
 30 <212> PRT  
 <213> unknown  
 <220>  
 35 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 33  
 40 Met Lys Ala Ile Gln Tyr Ala Arg Ile Gly Ala Glu Pro Glu Leu Thr  
 1 5 10 15  
 45 Glu Ile Pro Lys Pro Glu Pro Gly Pro Gly Glu Val Leu Leu Glu Val  
 20 25 30  
 Thr Ala Ala Gly Val Cys His Ser Asp Asp Phe Ile Met Ser Leu Pro  
 35 40 45  
 50 Glu Glu Gln Tyr Thr Tyr Gly Leu Pro Leu Thr Leu Gly His Glu Gly  
 50 55 60  
 55 Ala Gly Arg Val Ala Ala Val Gly Glu Gly Val Glu Gly Leu Asp Ile  
 65 70 75 80

Gly Thr Asn Val Val Val Tyr Gly Pro Trp Gly Cys Gly Ser Cys Trp  
                                     85                                    90                                    95

5 His Cys Ser Gln Gly Leu Glu Asn Tyr Cys Ser Arg Ala Lys Glu Leu  
                                     100                                    105                                    110

10 Gly Ile Asn Pro Pro Gly Leu Gly Ala Pro Gly Ala Leu Ala Glu Phe  
                                     115                                    120                                    125

15 Met Ile Val Asp Ser Pro Arg His Leu Val Pro Ile Gly Asp Leu Asp  
                                     130                                    135                                    140

20 Pro Val Lys Thr Val Pro Leu Thr Asp Ala Gly Leu Thr Pro Tyr His  
                                     145                                    150                                    155                                    160

Ala Ile Lys Arg Ser Leu Pro Lys Leu Arg Gly Gly Ala Tyr Ala Val  
                                     165                                    170                                    175

25 Val Ile Gly Thr Gly Gly Leu Gly His Val Ala Ile Gln Leu Leu Arg  
                                     180                                    185                                    190

30 His Leu Ser Ala Ala Thr Val Ile Ala Leu Asp Val Ser Ala Asp Lys  
                                     195                                    200                                    205

35 Leu Glu Leu Ala Thr Lys Val Gly Ala His Glu Val Val Leu Ser Asp  
                                     210                                    215                                    220

40 Lys Asp Ala Ala Glu Asn Val Arg Arg Ile Thr Gly Ser Gln Gly Ala  
                                     225                                    230                                    235                                    240

Ala Leu Val Leu Asp Phe Val Gly Tyr Gln Pro Thr Ile Asp Thr Ala  
                                     245                                    250                                    255

45 Met Ala Val Ala Gly Val Gly Ser Asp Val Thr Ile Val Gly Ile Gly  
                                     260                                    265                                    270

50 Asp Gly Gln Ala His Ala Lys Val Gly Phe Phe Gln Ser Pro Tyr Glu  
                                     275                                    280                                    285

55 Ala Ser Val Thr Val Pro Tyr Trp Gly Ala Arg Asn Glu Leu Ile Glu  
                                     290                                    295                                    300

Leu Ile Asp Leu Ala His Ala Gly Ile Phe Asp Ile Ala Val Glu Thr  
                                     305                                    310                                    315                                    320

5 Phe Ser Leu Asp Asn Gly Ala Glu Ala Tyr Arg Arg Leu Ala Ala Gly  
 325 330 335  
 10 Thr Leu Ser Gly Arg Ala Val Val Val Pro Gly Leu  
 340 345  
 15 <210> 34  
 <211> 348  
 <212> PRT  
 <213> unknown  
 <220>  
 20 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 34  
 25 Met Lys Ala Ile Gln Tyr Thr Arg Ile Gly Ala Glu Pro Glu Leu Thr  
 1 5 10 15  
 30 Glu Ile Pro Lys Pro Glu Pro Gly Pro Gly Glu Val Leu Leu Glu Val  
 20 25 30  
 35 Thr Ala Ala Gly Val Cys His Ser Asp Asp Phe Ile Met Ser Leu Pro  
 35 40 45  
 40 Glu Glu Gln Tyr Thr Tyr Gly Leu Pro Leu Thr Leu Gly His Glu Gly  
 50 55 60  
 45 Ala Gly Arg Val Ala Ala Val Gly Glu Gly Val Glu Gly Leu Asp Ile  
 65 70 75 80  
 50 Gly Thr Asn Val Val Val Tyr Gly Pro Trp Gly Cys Gly Ser Cys Trp  
 85 90 95  
 His Cys Ser Gln Gly Leu Glu Asn Tyr Cys Ser Arg Ala Lys Glu Leu  
 100 105 110

Gly Ile Asn Pro Pro Gly Leu Gly Ala Pro Gly Ala Leu Ala Glu Phe  
 115 120 125

5 Met Ile Val Asp Ser Pro Arg His Leu Val Pro Ile Gly Asp Leu Asp  
 130 135 140

10 Pro Val Lys Thr Val Pro Leu Thr Asp Ala Gly Leu Thr Pro Tyr His  
 145 150 155 160

15 Ala Ile Lys Arg Ser Leu Pro Lys Leu Arg Gly Gly Ala Tyr Ala Val  
 165 170 175

Val Ile Gly Thr Gly Gly Leu Gly His Val Ala Ile Gln Leu Leu Arg  
 180 185 190

20 His Leu Ser Ala Ala Thr Val Ile Ala Leu Asp Val Ser Ala Asp Lys  
 195 200 205

25 Leu Glu Leu Ala Thr Lys Val Gly Ala His Glu Val Val Leu Ser Asp  
 210 215 220

30 Lys Asp Ala Ala Glu Asn Val Arg Arg Ile Thr Gly Ser Gln Gly Ala  
 225 230 235 240

35 Ala Leu Val Leu Asp Phe Val Gly Tyr Gln Pro Thr Ile Asp Thr Ala  
 245 250 255

Met Ala Val Ala Gly Val Gly Ser Asp Val Thr Ile Val Gly Ile Gly  
 260 265 270

40 Asp Gly Gln Ala His Ala Lys Val Gly Phe Phe Gln Ser Pro Tyr Glu  
 275 280 285

45 Ala Ser Val Thr Val Pro Tyr Trp Gly Ala Arg Asn Glu Leu Ile Glu  
 290 295 300

50 Leu Ile Asp Leu Ala His Ala Gly Ile Phe Asp Ile Ala Val Glu Thr  
 305 310 315 320

55 Phe Ser Leu Asp Asn Gly Ala Glu Ala Tyr Arg Arg Leu Ala Ala Gly  
 325 330 335

Thr Leu Ser Gly Arg Ala Val Val Val Pro Gly Leu  
 340 345



<210> 35  
 <211> 488  
 5 <212> DNA  
 <213> unknown  
 10 <220>  
 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 15 ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
 ZF0050330= Bacillus; ZF0051303= Bacterium; ZF0051337=  
 Methylobionas; ZF0051321= Bacterium; ZF0050782= Lactobacillus  
 bulgaricus; ZF0050544= Phyllobacterium rubiacearum; ZF0002852=  
 Rhodococcus; ZF0050310= Arthrobacter paraffineus; ZF0002862=  
 20 Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0002031=  
 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0002437= Streptomyces; ZF0003712=  
 Micromonospora; ZF0003765= Streptomyces; ZF0051305= Bacterium;  
 ZF0003513= Actinomyces; ZF0050993= Kocuria; ZF0002018=  
 25 Streptomyces; ZF0003767= Actinomyces; ZF0002332= Streptomyces  
 diastatochromogenes; ZF0003768= Actinomyces; ZF0002379=  
 Streptomyces coelestis; ZF0002351= Nonomuraea roseoviolacea;  
 ZF0003769= Actinomyces;  
 <400> 35  
 30 gggccatggg gttgtggcaa ctgttggcac tgctcacaag gactcgagaa ctattgctct  
 60  
 cgcgcccaag aactcggaat caatcctccc ggtctcggcg caccggcgcg gttggccgag  
 120  
 35 ttcatgatcg tcgattctcc tcgccacctt gtcccgatcg gtgacctoga ccgggtcaag  
 180  
 acggtgccgc tgaccgacgc cggctctgacg ccgtatcacg cgatcaagcg ttctctgccg  
 40 240  
 aaacttcgcg gaggctcgta cgcggttgctc attggtaccg gcgggctcgg ccacgtcgcc  
 300  
 45 attcagctcc tccgtcacct ctccggcgtca acggtcatcg ctttggacgt gagcgccggac  
 360  
 aagctcgaac tggcaaccaa ggtaggcgct cacgaagtgg ttctgtccga caaggacgcg  
 420  
 50 gccgagaacg tccgcaagat cactggaagt caaggcgccg cactggttct cgacttcgtt  
 480  
 ggctacca  
 55 488  
 <210> 36

<211> 385  
 <212> DNA  
 5 <213> unknown  
 <220>  
 <221> source  
 10 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
 ZF0050330= Bacillus; ZF0051303= Bacterium; ZF0051337=  
 Methylobacter; ZF0051321= Bacterium; ZF0050782= Lactobacillus  
 15 bulgaricus; ZF0050544= Phyllobacterium rubiacearum; ZF0002852=  
 Rhodococcus; ZF0050310= Arthrobacter paraffineus; ZF0002862=  
 Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0002031=  
 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0002437= Streptomyces; ZF0003712=  
 20 Micromonospora; ZF0003765= Streptomyces; ZF0051305= Bacterium;  
 ZF0003513= Actinomyces; ZF0050993= Kocuria; ZF0002018=  
 Streptomyces; ZF0003767= Actinomyces; ZF0002332= Streptomyces  
 diastatochromogenes; ZF0003768= Actinomyces; ZF0002379=  
 Streptomyces coelestis; ZF0002351= Nonomuraea roseoviolacea;  
 ZF0003769= Actinomyces;  
 25 <400> 36  
 gggccatggg gttgtggcaa ctgttggcac tgctcacaag gactcgagaa ctattgctct  
 60  
 cgcgcccaag aactcggaat caatcctccc ggtctcggtg caccggcggc gttggccgag  
 30 120  
 ttcattgatcg tcgattctcc tcgccacctt gtcccgatcg gtgacctcga cccgggtcaag  
 180  
 35 acggtgccgc tgaccgacgc cgggtctgacg ccgtatcacg cgatcaagcg ttctctgccg  
 240  
 aaacttcgcg gaggtctgta cgcggttgctc attggtaccg gcgggctcgg ccacgtcacc  
 300  
 40 attcagctcc tccgtcacct ctccggcgga acggtcatcg ctttggacgt gagcgcgga  
 360  
 aagctcgaac tggcaaccaa ggtag  
 45 385  
 <210> 37  
 50 <211> 486  
 <212> DNA  
 <213> unknown  
 55 <220>  
 <221> source  
 <223> ZF0050286= Corynebacterium hoagii

<400> 37  
ggcccttggg gttgcggacg ttgctggcac tgcgcgcagg ggctcgagaa ctactgctcc  
60  
5 cgcgcaaggg aactcggcat cgccccaccc ggcttggggc cgccggggcg gatcgccgag  
120  
tacatgatcg tcgactcgcc gcgtcacctg gtcccgatcg gtgacctoga ccccgtcacg  
10 180  
acggtgccgc tgaccgacgc cgggctcacc cegtaccacg cgatcaaacg gtcgctcggc  
240  
15 aagctccgcg ccggctcgta cgcagtcgtg atcggcaccg gaggcctcgg acacgtcggc  
300  
atccagctgc tccgccacct gtcccctgca cgcatactcg ccctcgacgt caacgacgag  
360  
20 aagctcgcgt tcgcccgcga ggctcggcgcg caccgagaccg tgttgtcgaa cgccgacgcc  
420  
gccgcgaacg tccggaagat caccgggttcg gccgggtgccg cgctggctct agacttcgtc  
25 480  
ggctac  
486  
30  
<210> 38  
<211> 483  
35 <212> DNA  
<213> unknown  
<220>  
40  
<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*  
<400> 38  
45 ggcccatggg gctgtggcag ctgttggcac tgctcgcaag gactcgaaaa ctactgttct  
60  
cgggcaaaaag aactcggcat caatcctcct ggtctcgggtg caccggggcg gttggccgaa  
120  
50 ttcatgatcg tcgattcacc tcgccacctc gtcccgatcg gcgacctoga tccgggtcaag  
180  
acggtgccac tgaccgacgc cggctctgact cegtatcacg cgatcaagcg ttcactgccg  
55 240  
aaacttcgcg gtggcgcgta cgccgtcgtc atcgggtaccg gcgggtctcgg ccattgtcgcc  
300

```

atccaactcc tccgccacct ctcggcagca accgtcatcg cactcgacgt gagcgcggaac
360

aagctcgtac tggcaaccaa ggtaggcgct cacgaagtgg tcctgtccga caaggacgcg
5 420

gccgagaacg tccgcaggat caccggaagt cagggcgccg cactggttct tgacttcggt
480

10 ggc
483

<210> 39
15 <211> 210
<212> DNA
20 <213> unknown
<220>
<221> source
25 <223> ZF0004210= Actinomyces; ZF0004212= Actinomyces; ZF0004211=
Actinomyces; ZF0003860= Actinomyces; ZF0004218= Actinomyces;
ZF0003868= Actinomadura; ZF0004213= Actinomyces; ZF0003876=
Actinomyces; ZF0003866= Actinomyces; ZF0003864= Actinomyces;
ZF0003862= Actinomadura; ZF0003869= Actinomyces; ZF0003867=
30 Actinomadura; ZF0004216= Actinomyces; ZF0004235= Actinomyces;
ZF0004209= Actinomadura; ZF0004214= Actinomyces; ZF0003871=
Actinomyces; ZF0004063= Actinomadura; ZF0004052= Actinomadura;
ZF0006405= Streptomyces; ZF0003865= Actinomadura; ZF0004047=
Actinomadura; ZF0004070= Actinomyces; ZF0004085= Actinomyces;
35 ZF0004217= Actinomyces; ZF0004089= Actinomadura; ZF0004090=
Actinomadura; ZF0006138= Streptomyces; ZF0004236= Actinomadura;
ZF0051203= Bacterium;

<400> 39
40 ggaccgtggg gctgcggcac gtgcgtcaag tgccgccgagg gcaaggagaa ctactgcctg
60

cgcgccaagg aactcggcat cgccccgcc ggactcggct cgccccggcg catggccgag
45 120

tacatgatcg tcgacgaccc gcgccacctg gtgccgctcg gcggtctcga cccgggtccag
180

gccgtgccgc tcactgacgc gggcctgaca
50 210

<210> 40
55 <211> 282
<212> DNA
<213> unknown

```

<220>

<221> source

5 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
10 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
Actinoplanes philippinensis; ZF0002441= Streptomyces;  
ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
15 Actinomyces; ZF0002359= Actinoplanes lanthinogenes; ZF0002396=  
Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
20 Actinomyces; ZF0003535= Actinomyces;

<400> 40  
tgtcacaccg atcaccacat cgtcaccggc gcgaccccca tgccgtcggt cccgggtcatg  
60  
25 ggcggggcacg aggggttcggg cgtcatcacc aagctcggcc ctgaggtcaa gggactggag  
120  
gtcggcgacc acgtcggttct gtccttcatt ccggcttggtg gaacctgtcc ggcgtgttcg  
30 180  
gccgggcatc agaattctttg tgacctcggg atgggcctcc tcagcggcca agccatcagc  
240  
35 gacgggcacgt accggatcca ggctcgcggc gaaaacgtga tc  
282

<210> 41

40 <211> 276

<212> DNA

45 <213> unknown

<220>

<221> source

50 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
55 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
Actinoplanes philippinensis; ZF0002441= Streptomyces;  
ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
Micromonospora; ZF0004980= Streptomyces; ZF0004821=

Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396= Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces; ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087= Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium; ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538= Actinomyces; ZF0003535= Actinomyces;

5

<400> 41  
 tgcataaccg acgatcatgc tgtgaccggt gatctggcag tcccactccc cgtgatcggt  
 10 60  
 ggccacgaag gcgcgggcat agtggagaaa gtcggccccg gcgtgcgaga cgtcgaggta  
 120  
 ggcatcacg tcgtcctctc cttcattccc tcgtgtggac gctgccgttg gtgcgagtc  
 15 180  
 ggacagagca acctctgcga cctcggcgcc attctgatgg ccggcgacaca ggtcgacggg  
 240  
 20 acgtaccgcg cgacagctcg cgggcacgac gtcgga  
 276  
 <210> 42  
 25 <211> 276  
 <212> DNA  
 <213> unknown  
 30 <220>  
 <221> source  
 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 35 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031= Streptomyces;  
 ZF0002349= Streptomyces spectabilis; ZF0002434= Streptomyces;  
 ZF0050993= Kocuria; ZF0002018= Streptomyces; ZF0003767= Actinomyces;  
 40 ZF0003764= Streptomyces; ZF0002331= Actinoplanes philippinensis;  
 ZF0002441= Streptomyces; ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240= Bacterium;  
 ZF0002333= Rhodococcus erythropolis; ZF0003713= Micromonospora;  
 ZF0004980= Streptomyces; ZF0004821= Actinomyces;  
 45 ZF0002359= Actinoplanes ianthinogenes; ZF0002396= Actinoplanes;  
 ZF0003781= Actinomyces; ZF0003512= Actinomyces; ZF0006093= Streptomyces;  
 ZF0006103= Streptomyces; ZF0006087= Streptomyces; ZF0050446= Bacterium;  
 ZF0050445= Bacterium; ZF0006086= Streptomyces; ZF0002322= Rhodococcus;  
 ZF0003538= Actinomyces; ZF0003535= Actinomyces;  
 50

<400> 42  
 tgccatacag acgatcatgc tgtgaccggt gatctggcag tcccactccc cgtgatcggt  
 60  
 55 ggccacgaag gcgcgggcat agtggagaaa gtcggccccg gcgtgcgaga cgtcgaggta  
 120  
 ggcatcacg tcgtcctctc cttcattccc tcgtgtggac gctgccgttg gtgcgagtc  
 180

ggacagagca acctctgcga cctcggcgcc attctgatgg ccggcgacaca ggtcgacggg  
 240

5 acgtaccgcg cgacagctcg cgggcacgac gtcgga  
 276

10 <210> 43  
 <211> 276  
 <212> DNA

15 <213> unknown  
 <220>  
 <221> source  
 20 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
 ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
 ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
 Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
 Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
 25 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
 Actinoplanes philippinensis; ZF0002441= Streptomyces;  
 ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
 Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
 Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
 30 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
 Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
 ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
 Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
 ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
 35 Actinomyces; ZF0003535= Actinomyces;

<400> 43  
 tgtcactactg acgatcatgc tgtgaccggt gatctggcag tcccactccc cgtgatcggt  
 60

40 ggccacgaag gcgcgggcat agtggagaaa gtcggccccg gcgtgcgaga cgtcgaggta  
 120

45 ggcgatcacg tcgtcctctc cttcattccc tcgtgtggac gctgccgttg gtgcgcagtc  
 180

ggacagagca acctctgcga cctcggcgcc attctgatgg ccggcgacaca ggtcgacggg  
 240

50 acgtaccgcg cgacagctcg cgggcacgac gtcgga  
 276

55 <210> 44  
 <211> 276  
 <212> DNA

<213> unknown

<220>

5 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
10 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
Actinoplanes philippinensis; ZF0002441= Streptomyces;  
ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
Micromonospora; ZF0004980= Streptomyces; ZF0004821=  
15 Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396=  
Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces;  
ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087=  
Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium;  
ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538=  
20 Actinomyces; ZF0003535= Actinomyces;

<400> 44

tggtcacaccg acgatcatgc tgtgaccggt gatctggcag tccactccc cgtgatcggt  
60

25

ggccacgaag gcgcgggcat agtggagaaa gtcggccccg gcgtgcgaga cgtcgaggta  
120

30

ggcgatcacg tcgtcctctc cttcattccc tcgtgtggac gctgccgttg gtgcgcagtc  
180

ggacagagca acctctgcga cctcggcgcc attctgatgg ccggcgcacg ggtcgacggg  
240

35

acgtaccgcg cgacagctcg cgggcacgac gtcgga  
276

<210> 45

40

<211> 276

<212> DNA

45

<213> unknown

<220>

<221> source

50 <223> ZF0002326= Actinoplanes missouriensis; ZF0003505= Streptomyces;  
ZF0051321= Bacterium; ZF0050782= Lactobacillus bulgaricus;  
ZF0050544= Phyllobacterium rubiacearum; ZF0002031=  
Streptomyces; ZF0002349= Streptomyces spectabilis; ZF0002434=  
Streptomyces; ZF0050993= Kocuria; ZF0002018= Streptomyces;  
55 ZF0003767= Actinomyces; ZF0003764= Streptomyces; ZF0002331=  
Actinoplanes philippinensis; ZF0002441= Streptomyces;  
ZF0051307= Bacterium; ZF0051301= Bacterium; ZF0051240=  
Bacterium; ZF0002333= Rhodococcus erythropolis; ZF0003713=  
Micromonospora; ZF0004980= Streptomyces; ZF0004821=



Actinomyces; ZF0002359= Actinoplanes ianthinogenes; ZF0002396= Actinoplanes; ZF0003781= Actinomyces; ZF0003512= Actinomyces; ZF0006093= Streptomyces; ZF0006103= Streptomyces; ZF0006087= Streptomyces; ZF0050446= Bacterium; ZF0050445= Bacterium; ZF0006086= Streptomyces; ZF0002322= Rhodococcus; ZF0003538= Actinomyces; ZF0003535= Actinomyces;

5

<400> 45  
 10 tgtcacactg acgatcatgc tgtgaccggt gatctggcag tccactccc cgtgatcggg  
 60  
 gccacgaag gcgcgggcat agtggagaaa gtcggccccc gcgtgcgaga cgtcgaggta  
 120  
 15 ggcgatcacg tcgtcctctc cttcattccc tcgtgtggac gctgccgttg gtgcgcagtc  
 180  
 ggacagagca acctctgcga cctcggcgcc attctgatgg ccggcgacaca ggtcgacggg  
 240  
 20 acgtaccgcg cgacagctcg cgggcacgac gtcgga  
 276  
 <210> 46  
 25 <211> 279  
 <212> .DNA  
 <213> unknown  
 30 <220>  
 <221> source  
 <223> ZF0050310= Arthrobacter paraffineus  
 35 <400> 46  
 tgccacacag atctgttcac gaagtcgggtg ctaccggaaa ggctcggccc ctgcgtgttc  
 60  
 40 gggcacgaag gagcgggggt ggtcgaggcc gtcggctcgt cgatcgacag cattgcgccc  
 120  
 ggtgatcacg tgttgctgag ctaccgcagt tgcgggtgtg gcaggcagtg cctcagcggg  
 180  
 45 catcgggct actgcgaaag ctcacacggg ctcaacagct ctggcgacag caccgaaggc  
 240  
 tcgacgccgg tccggcgaag cggaactccg atacggtcc  
 50 279  
 <210> 47  
 55 <211> 279  
 <212> DNA  
 <213> unknown

<220>

<221> source

5 <223> ZF0002333= *Rhodococcus erythropolis*

<400> 47

tgatcatactg atctgttcac gaagacggtg ctaccggaaa agctcggggc ctgcgtgttc  
60

10 ggacacgaag gcgccggcgt cgtgcaagcc gttggctcgt cgatcgacaa catcgcggct  
120

ggatgatcacg tattgctgag ctaccgcagt tgcgggtgat gcaggcaatg tctcagcgac  
15 180

catcggggcgt actgcgaaag ctcacacggg ctcaacagct ctggcgacag caccgacggc  
240

20 tgcacgccgg tccggcgaaa cggaactccg atacgggtcc  
279

<210> 48

25 <211> 360

<212> DNA

<213> unknown

30 <220>

<221> source

35 <223> ZF0051303= *Bacterium*; ZF0051337= *Methylobacter*; ZF0002862=  
*Streptomyces clavuligerus*; ZF0050292= *Bacterium*; ZF0051305=  
*Bacterium*; ZF0003513= *Actinomyces*; ZF0002351= *Nonomuraea*  
*roseoviolacea*; ZF0003769= *Actinomyces*; ZF0002017= *Streptomyces*;  
ZF0051306= *Bacterium*; ZF0002016= *Streptomyces*; ZF0003504=  
40 *Actinomyces*; ZF0006073= *Streptomyces*; ZF0003770= *Actinomyces*;  
ZF0002352= *Actinoplanes italicus*; ZF0002378= *Streptomyces*  
*aureomonopodiales*; ZF0006089= *Streptomyces*; ZF0006106=  
*Streptomyces*; ZF0051325= *Bacterium*; ZF0006108= *Streptomyces*;  
ZF0002440= *Streptomyces*; ZF0051302= *Bacterium*; ZF0003532=  
45 *Actinomyces*; ZF0003548= *Nocardiaform*;

<400> 48

tgccgggacgg accgcgagat cgcctcgggc atctacgggt gggcgccgcc gggacgcgaa  
60

50 cacctcgtcc tcggggcacga atcgctgggc cgagtacgca ccgcgcccga cggcagcgg  
120

ttcgccgccg gtgatctcgt cgtcgggata gtgcgcaggc ccgatccggt gccgtgcccc  
180

55 gcgtgtgcgc acggtgagtt cgacatgtgc cgcaacgggtg agtacgtcga gcgcgggata  
240

aagcagatcg acgggtacgg gtcgacgtcg tgggtgggtgg acgccgacta cacgggtcaag  
300

5 ctggaccccg cgctcaccca ggtgggtgtg ctgatggaac cgacgacggt gcttggccaa  
360

<210> 49

10 <211> 421

<212> DNA

<213> unknown

15 <220>

<221> source

20 <223> ZF0051303= Bacterium; ZF0051337= Methylomonas; ZF0002862=  
Streptomyces clavuligerus; ZF0050292= Bacterium; ZF0051305=  
Bacterium; ZF0003513= Actinomyces; ZF0002351= Nonomuraea  
roseoviolacea; ZF0003769= Actinomyces; ZF0002017= Streptomyces;  
ZF0051306= Bacterium; ZF0002016= Streptomyces; ZF0003504=  
Actinomyces; ZF0006073= Streptomyces; ZF0003770= Actinomyces;  
25 ZF0002352= Actinoplanes italicus; ZF0002378= Streptomyces  
aureomonopodiales; ZF0006089= Streptomyces; ZF0006106=  
Streptomyces; ZF0051325= Bacterium; ZF0006108= Streptomyces;  
ZF0002440= Streptomyces; ZF0051302= Bacterium; ZF0003532=  
Actinomyces; ZF0003548= Nocardiaform;

30 <400> 49  
tgtggtaccg acctgcacat ccggctcctgg gacggatggg cgcagaagac catcgccacc  
60

35 ccgctcacgc tcggccacga gttcgtcggc gaggtcgtcg agaccggccg cgacgtgacc  
120

gacatccagg tcggcgacct ggtcagcggc gagggccacc tggctctgcgg caagtgccgc  
180

40 aactgcctgg ccggccgccg tcacctgtgc cgcgcgaccg tcggcctcgg tgcgggccgt  
240

45 gacggcgctt tcgccgagta cgtgggtgctg cccgcctcca acgtgtgggt gcaccgggtg  
300

ccggctgacc tcgacgtcgc cgcgatcttc gacccgttcg gcaacgcggt gcacaccgcg  
360

50 ctctccttcc cgctcgtcgg cgaggacgtg ctggtcaccg gtgccgggtac catcggcata  
420

t  
421

55 <210> 50

<211> 414

<212> DNA  
 <213> unknown  
 5 <220>  
 <221> source  
 10 <223> ZF0050197= Pseudomonas oleovorans; ZF0050294= Rhodococcus;  
 ZF0050330= Bacillus, ZF0002852= Rhodococcus; ZF0050310=  
 Arthrobacter paraffineus; ZF0002437= Streptomyces; ZF0003712=  
 Micromonospora; ZF0003765= Streptomyces; ZF0002332=  
 Streptomyces diatsatochromogenes; ZF0003768= Actinomyces;  
 15 ZF0002379= Streptomyces coelestis; ZF0002443= Streptomyces;  
 ZF0002442= Streptomyces; ZF0002436= Streptomyces; ZF0050994=  
 Bacterium; ZF0050992= Bacterium; ZF0050442= Bacterium;  
 ZF0002049= Streptomyces; ZF0006069= Streptomyces; ZF0006075=  
 Streptomyces; ZF0004724= Nocardiaform; ZF0002392= Actinoplanes  
 nipponensis; ZF0002356= Actinoplanes brasiliensis; ZF0003501=  
 20 Actinomyces; ZF0051322= Bacterium; ZF0006078= Streptomyces;  
 ZF0006092= Streptomyces; ZF0006090= Streptomyces; ZF0006084=  
 Streptomyces; ZF0006068= Streptomyces; ZF0050284= Rhodococcus;  
 ZF0050028= Agrobacterium tumefaciens; ZF0003540= Actinomyces;  
 25 ZF0003528= Actinomyces; ZF0003529= Actinomyces;  
 <400> 50  
 ggcctgacga tcggccatga accggtgggc gtcacgaaa agctgggcag cgccgtgacg  
 60  
 30 gggtaccgag agggccaacg cgtgatcgcc ggccgcatct gccccaactt caattcgat  
 120  
 gccgcgagg atggcgccc gtcgcaggat ggccagctacc tggggccag cgccgcatgc  
 180  
 35 ggctgccatg gataccgggc caccggcggc tggcgctttg gcaacatcat cgatggcgcc  
 240  
 caggccgaat acctgctggt tcccgatgag cagggaatc tggcgccggt tccggacaac  
 40 300  
 ctgagcgatg aacaggtgct gatgtgccc gacatcatgt ccaccggctt caaaggcgca  
 360  
 45 gagaacgcac acatccgcat cggcgacag gtggcggtat ttgcgcaggg acca  
 414  
 <210> 51  
 50 <211> 432  
 <212> DNA  
 55 <213> unknown  
 <220>  
 <221> source

<223> ZF0050197= *Pseudomonas oleovorans*; ZF0050294= *Rhodococcus*;  
 ZF0050330= *Bacillus*, ZF0002852= *Rhodococcus*; ZF0050310=  
*Arthrobacter paraffineus*; ZF0002437= *Streptomyces*; ZF0003712=  
*Micromonospora*; ZF0003765= *Streptomyces*; ZF0002332=  
 5 *Streptomyces diatsatochromogenes*; ZF0003768= *Actinomyces*;  
 ZF0002379= *Streptomyces coelestis*; ZF0002443= *Streptomyces*;  
 ZF0002442= *Streptomyces*; ZF0002436= *Streptomyces*; ZF0050994=  
*Bacterium*; ZF0050992= *Bacterium*; ZF0050442= *Bacterium*;  
 10 ZF0002049= *Streptomyces*; ZF0006069= *Streptomyces*; ZF0006075=  
*Streptomyces*; ZF0004724= *Nocardiaform*; ZF0002392= *Actinoplanes*  
*nipponensis*; ZF0002356= *Actinoplanes brasiliensis*; ZF0003501=  
*Actinomyces*; ZF0051322= *Bacterium*; ZF0006078= *Streptomyces*;  
 ZF0006092= *Streptomyces*; ZF0006090= *Streptomyces*; ZF0006084=  
*Streptomyces*; ZF0006068= *Streptomyces*; ZF0050284= *Rhodococcus*;  
 15 ZF0050028= *Agrobacterium tumefaciens*; ZF0003540= *Actinomyces*;  
 ZF0003528= *Actinomyces*; ZF0003529= *Actinomyces*;  
  
 <400> 51  
 20 tgcgggacgg acctgcacat cctcggaggt gacgtcccg aggtgaccga cgggcgaatc  
 60  
 ctgggccacg aggcgctcgg gaccgtgggtc gaggtgggcg acggcgatca gacactcggc  
 120  
 25 ccgggcgatc gcgtgctcgt ctggtgtgtc accgcatgcg gtacgtgccg gttctgccgc  
 180  
 gagagccgct acgggcgaatg cctcggaggc ggcggctgga tcctcggaca cctgatcgac  
 240  
 30 ggcaaccagg ccgaactcgt ccgagttcgg tacgccgaca attcgaccca ccgcatcccc  
 300  
 gacggtgtga gcgacgagca gatgctcatg ctgcgcgaca tcctgcccac ctctacgag  
 35 360  
 gtcggtgttc tcaacggctg tctccggcgg ggcgacgtcg tcgtcatcat cggggccgac  
 420  
 40 gatcggcctc tt  
 432  
  
 <210> 52  
 45 <211> 220  
 <212> DNA  
 50 <213> unknown  
 <220>  
 <221> source  
 55 <223> ZF0050310= *Arthrobacter paraffineus*  
  
 <400> 52  
 cgtcgacgtc gtcgtcgaca acgcgggatt ccgaacacac ggggcattcg tggacgaaga  
 60

tcacgagcgc gtcacgtccg agattcagct caacatcgcc acgctggtcg agctgacaca  
120

5 cacattcccc cccgaccttc tcaccggccg cggagcactg gtcaacatcg ccagcacagc  
180

gtcgttccag cccgacaccgg gcatggccgt ctactgcgct  
220

10

<210> 53

<211> 226

15 <212> DNA

<213> unknown

20 <220>

<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*

25 <400> 53  
cgtcgacgtc gtcgtccaca acgccggatt cggaacacac ggggcattcg tggacgaaga  
60

30 tctcgagcgc gtcacgtccg agattcagct caacatcgcc acgctggtcg agctgacaca  
120

cacattcctg cccgaccttc tcaccggccg cggagcactg gtcaacatcg ccagcacagc  
180

35 gtcgttccag cccgacaccgg gcatggccgt ctactgcgcc accaag  
226

<210> 54

40 <211> 237

<212> DNA

45 <213> unknown

<220>

<221> source  
50 <223> ZF0003535= *Actinomyces*

<400> 54  
cgtgtcgacg tcgtggtgca caatgctgcg atcactcaaa aggccacttt tcgcgacatt  
60

55 acccccgccg attttgagcg catcctgcgg gtgaacctga ccggcgctct caacctgagc  
120

caagccgtca ttcccttgat gattcagcgc ggcggaggaa gcatcgtctc gatttcctcg  
180

5 ctgtcggcgc agaacggcgg ggggatcttc ggcggcgccc actattgcgc aaccaag  
237

<210> 55

10 <211> 229

<212> DNA

<213> unknown

15 <220>

<221> source

<223> ZF0003535= Actinomyces

20 <400> 55  
cgtcgacgtc gtcgtcgaca acgccggtct ggcactgggc acggcccccg cgccgcaggt  
60

25 gccgctaaag gactggcaga ccatggtgaa caccaacatc accggtctac tgaacatcac  
120

ccaccatctc ctgccgacac tgatcgaccg taaaggatc gtcgtcaacc ttctgtctgt  
180

30 tgccgcgcac tatccctata cgggcggcaa tgtatactgc gcctccaag  
229

35 <210> 56

<211> 216

<212> DNA

40 <213> unknown

<220>

45 <221> source

<223> ZF0050310= Arthrobacter paraffineus

<400> 56

50 caggggatcg gatacgccac cgccaagcgg ctgatcagcc tgggtgcgac ggtcgcgac  
60

ggcgacatcg acgaagccac tctcgcgcga gcagccaagg atttgggcat ccgcacgttc  
120

55 gggcgctctg acgtcaccga ccccgctctg ttcttcgact tctcgacac cgtcgaaggt  
180

gaactcggcc cgatcgacgt gctgatcaac aacgcg  
216

<210> 57  
5 <211> 225  
<212> DNA  
<213> unknown  
10 <220>  
<221> source  
<223> ZF0080310= *Arthrobacter paraffineus*  
15 <400> 57  
cagcggatcg ggctcgaaat tgcgcgacac ttcacaaagg aaggcgcgac cgtcggttctc  
60  
20 ggcgacatca acgaaaccgt gggaacggct gcggtcgccg aactcgggtg agagtcggtc  
120  
gcctgtttcg cttcctgcga cgtgcgtgac tccggacagg tcgaggccat gtcgatctg  
180  
25 gccgaaagcg ctttcggtcc agtcgatgtc atgatgaaca acgcg  
225  
<210> 58  
30 <211> 216  
<212> DNA  
35 <213> unknown  
<220>  
<221> source  
40 <223> ZF0080310= *Arthrobacter paraffineus*  
<400> 58  
caggggatcg gctaccagac cgcaaggag ctgatccgac gaggtcaccg cgtggccatc  
60  
45 ggcgacatcg acgaggcacg tgctaaggag accgccgccg aactgggggt taaggttgtc  
120  
accgcctcg atgtaccga ccctgactcg ttcaaagact ttctcgacct agtcgagga  
50 180  
gacctcggcc ccctcgacgt gctgatcaac aacgcg  
216  
55 <210> 59  
<211> 222



<212> DNA  
<213> unknown

5 <220>  
<221> source  
<223> ZF0080310= *Arthrobacter paraffineus*

10 <400> 59  
gggatcgggc tcgaaattgc gcgcaccttc atcaaggaag gcgcgaccgt cgttctcggc  
60  
gacatcaacg aaaccgtggg aacggctgcg gtcgccgaac tcgggtggaga gtccggtcgcc  
15 120  
cgtttcgctt cctgcgacgt gcgtgactcc ggacaggctg aggccatgct cgatctggcc  
180  
gaaagcgctt tcgggtccagt cgatgtcatc gtgaacaacg cg  
20 222  
<210> 60  
<211> 222

25 <212> DNA  
<213> unknown

30 <220>  
<221> source  
<223> ZF0080310= *Arthrobacter paraffineus*

35 <400> 60  
atcgggctcg aaattgcgcg caccttcacg aaggaaggcg cgaccgtcgt tctcggcgac  
60  
atcaacgaaa ccgtgggaac ggctgcggtc ggcgaactcg gtggagagtc ggtcgcccgt  
40 120  
ttcgcttcct gcgacgtgcg tgactccgga caggctcgagg ccatgctcga tctggccgaa  
180

45 agcgctttcg gtccagtcga tgtcatggtc aacaacgccg gc  
222

50 <210> 61  
<211> 186  
<212> DNA

55 <213> unknown  
<220>  
<221> source

<223> ZF0002333= Rhodococcus erythropolis

<400> 61

5 gtgccggtcg cggtcgtgga ccttcacatc gaaagtgcaa aggagaccgt cgcacttatc  
60

gaatcgcagt acggcacacc cgcgctcgcc cttgaggccg atgtgcgcga ccgcgcgcgc  
120

10 gtgagcgccg ctttcgaagc caccgtcgcc gaatggggac gcttcgacta cctcgtcaac  
180

aacgcc  
186

15

<210> 62

<211> 222

20

<212> DNA

<213> unknown

<220>

25

<221> source

<223> ZF0002333= Rhodococcus erythropolis

<400> 62

30 ctcggccgtg aaatcgctct caagctcgct tccgaaggcg cctcggtagt ggtcaacgac  
60

ctcgatcccc aacctgccgc tcagaccgag cgcgatatca aagccacagg tggacaggct  
120

35

gtctcgtgcg tcggctccgt tgccgaggac gggttcgccg aacgcttcgt gaacactgcc  
180

gtcgaatcat tcggcggact cgacgtcatg gtgaacaacg cg  
222

40

<210> 63

45

<211> 231

<212> DNA

<213> unknown

50

<220>

<221> source

<223> ZF0002333= Rhodococcus erythropolis

55

<400> 63

gcggggctcg gagtggaatt cgctcaccgc ttcgccgctc gcggtgcaaa tctggttctc  
60

gtcgccaggc gggcagatcg cctcgaagcc ctcgctaccg aactccgcgt cgcccacggc  
120

5 atcacagtca cagttctgcc tgccgacctg gggcgcccg gcgtcggcgc aacactgcac  
180

caggagctga caagccgtgg catcacgctc acctcgctga tcaacaacgc c  
231

10 <210> 64

<211> 216

15 <212> DNA

<213> unknown

<220>

20 <221> source

<223> ZF0003535= Actinomyces

<400> 64

25 ccagcggacg gctatcagac agcgaaggag ttgattcgac gaggccaccg ggtcgccatc  
60

gtcgacatcg acgaggcacg tgcgaagggg gccgcccgcg aactcggggg gaaggctcgc  
120

30 acccgactcg acgtcacoga acctgactcg ttcacaacgt ttctggacct ggtcgaacgt  
180

gaactcggac ccctcgacat cctgggtcaac aacgcg  
216

35 <210> 65

40 <211> 201

<212> DNA

<213> unknown

45 <220>

<221> source

<223> ZF0050310= Arthrobacter paraffineus

50 <400> 65

gccacggacg gtgcccgcgt cgcggctcgtg gaccttcaca tcgaaagtgc agaggagacc  
60

55 gtcgcaetta tcgaatcgca gtacggcaca cccgcgctcg cccttgaggc cgatgtgcgc  
120

gaccgcgccg ccgtgagcgc cgctttcgaa gccaccgtcg ccgaatgggg acgcttcgac  
180

tacctcgtca acaacgccgg c  
201

5  
<210> 66  
<211> 201  
  
<212> DNA  
10  
<213> unknown  
  
<220>  
  
15  
<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*  
  
<400> 66  
20 gccgcggacg gtgcccgcgt cgcggtcgtg gaccttcaca tcgaaagtgc aaaggagacc  
60  
  
gtcgcaactta tcgaatcgca gtacggcaca cccgcgctcg cccttgaggc cgatgtgcgc  
120  
  
25 gaccgcgccg ccgtgagcgc cgcttttcgaa gccaccgtcg ccgaatgggg acgcttcgac  
180  
  
tacctcgtca acaacgccgg c  
201

30  
  
<210> 67  
  
<211> 1047  
35  
<212> DNA  
  
<213> unknown  
  
<220>  
  
40  
<221> source  
<223> ZF0050310= *Arthrobacter paraffineus*  
  
45  
<400> 67  
atgaaggcaa tccagtacgc gagaatcggc gcagaaccgc aactcacgga gattcccaaa  
60  
  
cccagagccc gtccaggtga agtgcctctg gaagtcaccg ctgccggcgt ctgccactcg  
50  
120  
  
gacgacttca tcatgagcct gccgaagag cagtacacct acggccttcc tctcacgctc  
180  
  
55  
ggccacgaag gcgccggccg ggctgccgcc gtcggcgagg gcgtcgaagg actcgacatc  
240  
  
ggaaccaatg tcgtcgtcta cggaccctgg ggctgtggca gctgttggca ctgctcgcaa  
300

ggactcgaaa actactgttc tcgggcaaaa gaactcggca tcaatcctcc tggctcgggt  
 360  
 5 gcacccggcg cgttggccga attcatgata gtcgattcac ctcgccacct cgtcccgatc  
 420  
 ggcgacctcg atccgggtcaa gacgggtgcca ctgaccgacg ccgggtctgac tccgtatcac  
 480  
 10 gcgatcaagc gttcactgcc gaaacttcgc ggtggcgcggt acgccgtcgt catcgggtacc  
 540  
 ggcgggtctcg gccatgtcgc catccaactc ctccgccacc tctcggcage aaccgtcacc  
 600  
 15 gcactcgacg tgagcgcgga caagctcgaa ctggcaacca aggtaggcgc tcacgaagtg  
 660  
 gtcctgtccg acaaggacgc ggccgagAAC gtccgcagga tcaccggaag tcagggcgcc  
 720  
 20 gcactgggttc tcgacttcgt cggctatcag cccaccatcg acaccgcgat ggtgtctgcc  
 780  
 ggcgtcggat cggacgtcac gatcgtcggg atcggcgacg ggcaggccca tgccaaagtc  
 840  
 25 gggttcttcc aaagtcttta cgaggcttct gtgacagttc cgtactgggg tgcccgcaac  
 900  
 30 gagctgatcg aattgatcga cctggcgcac gccggcatct tcgacatcgc ggtggagacc  
 960  
 ttcagtctcg acaacggcgc cgaagcgtat cgacgactgg ccgccggaac gtcagcggc  
 1020  
 35 cgcgcgggttg tggtccttgg tctgtag  
 1047  
 40  
 <210> 68  
 <211> 1047  
 45 <212> DNA  
 <213> unknown  
 <220>  
 50 <221> source  
 <223> ZF0050310= *Arthrobacter paraffineus*  
 <400> 68  
 55 atgaaggcaa tccagtacac gagaatcggc gcagaacccg aactcacgga gattcccaaa  
 60  
 cccgagcccc gtccagggtga agtgctcctg gaagtcaccg ctgccggcgt ctgccactcg  
 120

gacgacttca tcatgagcct gcccgagag cagtacacct acggccttcc tctcacgctc  
180

5 ggcacgaag gcgccggccg ggtcgccgcc gtcggcgagg gcgtcgaagg actcgacatc  
240

ggaaccaatg tcgtcgtcta cggaccctgg ggctgtggca gctgttggca ctgctcgcaa  
300

10 ggactcgaaa actactgttc tcgggcacaaa gaactcggca tcaatcctcc tgggtctcgg  
360

gcacccggcg cgttggccga attcatgata gtcgattcac ctcgccacct cgtcccgatc  
420

15 ggcgacctcg atccggtcaa gacggtgcca ctgaccgacg ccggtctgac tccgtatcac  
480

20 gcgatcaagc gttcactgcc gaaacttcgc ggtggcgcggt acgccgtcgt catcggtacc  
540  
ggcgggtctcg gccatgtcgc catccaactc ctccgccacc tctcggcagc aaccgtcatc  
600

25 gcactcgacg tgagcgcgga caagctcgaa ctggcaacca aggtaggcgc tcacgaagtg  
660

gtcctgtccg acaaggacgc ggccgagAAC gtccgcagga tcaccggaag tcagggcgcc  
720

30 gcactggttc tcgacttcgt cggctatcag cccaccatcg acaccgcgat ggctgtcgcc  
780

ggcgtcggat cggacgtcac gatcgtcggg atcggcgacg ggcaggccca tgccaaagtc  
840

35 gggttcttcc aaagtctta cgaggcttct gtgacagttc cgtactgggg tgcccgcaac  
900

40 gagctgatcg aattgatcga cctggcgcac gccggcatct tcgacatcgc ggtggagacc  
960

ttcagtctcg acaacggcgc cgaagcgtat cgacgactgg ccgccggaac gtcagcggc  
1020

45 cgcgcgggtg tggtccctgg tctgtag

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☒ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**